The Secret Rites of Initiation into the Awesome Mysteries of Error Analysis & Lab Reports

J.B. Gaffney

Version of 13 August 2014

© 2001-2014 John Gaffney. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/ or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

A visitor to the Royal Tyrell Museum was admiring a tyrannosaurus fossil and asked a nearby museum employee how old it was.

"That skeleton's sixty-five million and three years, two months, and eighteen days old," the employee replied.

"How can you know it that well?" she asked.

"Well, when I started working here, I asked a scientist the exact same question, and he said it was sixty-five million years old — and that was three years, two months, and eighteen days ago."

(Author unknown)

Contents

1 Error Analysis **3**
- 1.1 *De Errorum Natura* . . . 3
- 1.2 Random & Systematic Error . . . 3
- 1.3 Absolute & Relative Error . . . 4
- 1.4 A Preliminary Example of Error Analysis . . . 5
- 1.5 Estimation of Error . . . 6
- 1.6 Statistical Estimation of Error . . . 7
- 1.7 Error Propagation . . . 9
- 1.8 Mean, Standard Deviation, Standard Error, & Confidence Levels 13
- 1.9 A More Complete Example of Error Analysis . . . 17
- 1.10 The Many Guises of Satan . . . 21
- 1.11 Exercises . . . 24

2 Fitting Curves **28**
- 2.1 The Least-Squares Fit . . . 28
- 2.2 Errors in Curve Fitting . . . 31
  - 2.2.1 The Correlation Coefficient . . . 31
  - 2.2.2 χ^2 & Standard Error . . . 33
- 2.3 Error Bars . . . 37

3 The Lab Report **38**
- 3.1 The Lab Notebook . . . 38
- 3.2 Feel Free to Use the First Person . . . 38
- 3.3 . . . But Do Speak English . . . 38
- 3.4 Anatomy of a Lab Report . . . 39
  - 3.4.1 The Title, &c. . . . 39
  - 3.4.2 The Abstract . . . 39
  - 3.4.3 The Procedure . . . 41
  - 3.4.4 Data & Tables . . . 41
  - 3.4.5 Calculations . . . 43
  - 3.4.6 Graphs . . . 43
  - 3.4.7 Error Analysis . . . 45
  - 3.4.8 Discussion . . . 45
  - 3.4.9 Conclusion . . . 47
  - 3.4.10 Putting the Pieces Together . . . 48
- 3.5 Outline of a Sample Lab Report . . . 48

A Examples of Abstracts **50**

B Proof of the Central Limit Theorem **56**

C The Standard Deviation Ain't the Average Distance from the Mean **65**

1 Error Analysis

1.1 *De Errorum Natura*

Unless they get therapy, perfectionists are doomed to live very unhappy lives. The world is a very imperfect place. Indeed, in the lab it may often seem to you that absolutely nothing goes right. What we want now is a quantitative means of measuring the unhappiness of perfectionists and your frustration in the lab.

In the real world[1] the purpose of an experiment is to determine the value of a physical quantity: you make a set of measurements, and from these you calculate a result for the physical quantity in question. For a pendulum, for example, the acceleration g due to gravity turns out to be given in terms of the length ℓ of the pendulum and the period T of its swings by

$$g = \frac{4\pi^2 \ell}{T^2}$$

By measuring the length of the pendulum and timing its swings, we can thus obtain a result for the acceleration due to gravity.

In a perfect world, our measured values of ℓ and T would be exact, and from them we would obtain an exact result for g. In our imperfect world, however, no measurement, no matter how carefully made or how precise the measuring device, is perfect, and consequently neither is any result derived from measurements. Experimental error is the limit on the accuracy of a measurement or result,[2] and error analysis is the two-part process of quantitatively ascertaining the errors in measurements and then, by a method known as error propagation, determining the corresponding error in a result derived from those measurements. It is only through such an error analysis that the validity and import of a result can be rigorously assessed: the smaller the error, the more accurate the result.[3]

The estimation, propagation, and interpretation of errors is fairly involved and will be discussed and developed in stages in §1.4 through §1.8. For now, the point is simply that there are limits to the accuracy of measurements and consequently to results derived from those measurements.

1.2 Random & Systematic Error

Experimental errors fall into two basic categories: random and systematic. Random errors are unbiased errors that are as likely to yield too high a value

[1] As opposed to the world in which you are now living. In your present pedagogical world, labs sometimes serve simply as demonstrations of phenomena discussed in class.

[2] Please burn this definition into the back of your skull. Failure to grasp this simple point is the cause of no end of mischief. And by mischief we don't mean joy-buzzer handshakes or leaving tacks on the teacher's chair; we mean weeping and gnashing of teeth. Your teeth. So do yourself a favor and make this definition the focus of your next meditation.

[3] Or, for the pessimists among you, the greater the error, the more inaccurate the result.

for a measurement as too low; systematic errors are biased errors that tend to go in one direction.

Suppose, as an admittedly very artificial and highly bogus example, but one that will illustrate the point, we have an object known to be revolving in a circle at a perfectly constant speed. If you use a stopwatch to make several measurements of how long it takes the object to complete one revolution, these time measurements will likely differ slightly from each other, even though the time interval should in principle be exactly the same: there will be variations in your judgement of when a full circle has been completed and in your reaction times when you start and stop the watch.[4,5] But as long as the stopwatch is properly calibrated, the time measurements will be as likely to be slightly above as slightly below the actual value of the time for one revolution, and the error in your time measurements is therefore random error.

Such random errors can, precisely because of their symmetry, be minimized by averaging over a set of repeated measurements; the deviations on the high and low side of the true value will tend to cancel each other. The more numerous the measurements, the closer their average will be to the true value.[6]

If the stopwatch were known to run fast,[7] this would introduce a systematic error into the time measurements: a watch that runs too fast will indicate that more time has elapsed than actually has, so that your recorded values for the time intervals will be consistently too high. And note that when there is systematic error in the time measurements, there is still random error as well; the random errors are merely superposed on top of the systematic error. Nobody ever said life was simple.

1.3 Absolute & Relative Error

There are two ways of expressing error quantitatively. The simplest way, known as absolute error, is to specify the magnitude of the inaccuracy directly. For example, if you obtain 4.52 cm when you measure the length of an object with a Vernier caliper accurate to 0.01 cm, the 0.01 cm — the amount by which your measurement of 4.52 cm might be higher or lower than the true length

[4] Nor do stopwatches measure time intervals continuously: each watch has some finite resolution — whether a tenth or a hundredth of a second or what-have-you — to which it rounds or truncates times. The granularity of the times recorded by the stopwatch is, however, beside our present point.

[5] You might object that it is boneheaded to time single revolutions: since all of the error in the time measurement occurs at the beginning and end of the interval, when you start and stop the stopwatch, measuring the total time for multiple revolutions and then dividing this total time by the number of revolutions will yield a time per revolution with proportionally smaller error. While this is true, it is, again, beside our present point.

[6] Just how much closer we will see later, when we discuss what is known as the standard deviation of the mean in §1.8.

[7] How would you know that the stopwatch was running fast? Unless it were running as absurdly fast as the watch in Mark Twain's *My Watch*, you could ascertain that it was miscalibrated only by comparing it to a watch known to be correctly calibrated.

— is the absolute error in the length measurement, and you would quote your measured value for the length as 4.52 ± 0.01 cm, meaning that the true length could be anywhere from 4.51 cm to 4.53 cm.

Alternatively, you can express the error in a value as a fraction of that value. This fractional reckoning is called relative error and is often expressed as a percentage. For the above length measurement, the relative error is

$$\frac{0.01}{4.52} \times 100 = 0.2\%$$

and you could quote your result for the measurement in terms of this relative error as $4.52\,\text{cm} \pm 0.2\%$.

1.4 A Preliminary Example of Error Analysis

Suppose you are calculating the density of a sample of lead from measurements of its mass and volume. You measure the mass to be 16.54 g on a scale accurate to 0.01 g, and you determine the volume, rather crudely, to be $1.2 \pm 0.5\ \text{cm}^3$ by submerging the sample in a graduated cylinder.[8] Your result for the density is thus[9]

$$\text{density} = \frac{\text{mass}}{\text{volume}} = \frac{16.54}{1.2} = 14\ \text{g/cm}^3$$

The expected value for the density — that is, the "book" value — is $11.25\ \text{g/cm}^3$, from which your result differs by

$$\frac{14 - 11.25}{11.25} = 24\%.$$

Looking at the 24% discrepancy, you might leap to the conclusion that your results are hopelessly wrong and that you made some horrible mistake in the lab. You might even get all emotional about it — angry about having wasted all that time or about having to take the stupid course to begin with. Tears of despair might follow oaths and imprecations.

This would be where a proper error analysis comes in.

To determine the density, you divided the mass by the volume. As we will see in §1.7, the relative error $\epsilon_{\text{density}}$ in the density is therefore given by

$$\epsilon_{\text{density}} = \sqrt{\epsilon_{\text{mass}}^2 + \epsilon_{\text{volume}}^2}$$

where ϵ_{mass} and ϵ_{volume} are the relative errors in the mass and volume, respectively:

$$\epsilon_{\text{mass}} = \frac{0.01}{16.54} \times 100 = 0.060\% \qquad \epsilon_{\text{volume}} = \frac{0.5}{1.2} \times 100 = 42\%.$$

[8] This is yet another respect in which physicists are superior to chemists — physicists don't use graduated cylinders.

[9] We assume, here and throughout, that you are already familiar with significant figures.

The relative error in the density thus works out to

$$\epsilon_{\text{density}} = \sqrt{\epsilon_{\text{mass}}^2 + \epsilon_{\text{volume}}^2} = \sqrt{0.060^2 + 42^2} \approx 42\%$$

Since 42% of 14 is 6, your result for the density is $14 \pm 6\ \text{g/cm}^3$. So although you obtained $14\ \text{g/cm}^3$ for the density, you can only be sure that it lies somewhere between 8 and $20\ \text{g/cm}^3$.

The above error analysis rigorously and quantitatively justifies several conclusions:

- Your result for the density in fact agrees, to well within experimental limits, with the book value: your error analysis indicates that your density result could, because of the inaccuracies in your measurements, be off by fully 42%, a margin within which 24% falls quite comfortably.
- The vast majority of the overall 42% error was due to the error in the volume measurement.
- To improve your result for the density, you would therefore do best to focus on finding a more accurate means of determining the volume of the sample.

To the extent that you consider a range of $\pm 42\%$ unacceptably wide, you might also, somewhat more subjectively, consider your results to be horrible. Such conclusions depend, however, on context: it might be that your results really are awful, or it might be that the required measurements are very difficult to make accurately with existing technology and techniques. Either way, it is extremely important to recognize that there is a difference between results being bad and being incorrect: your result of $14 \pm 6\ \text{g/cm}^3$ for the density of lead, however horrible in the sense of inaccurate, is perfectly correct; for your density result to be incorrect you would have to have made some sort of mistake, such as calculating the density from the mass and volume of two different samples of lead or forgetting to tare the scale before weighing the sample.

1.5 Estimation of Error

The estimation of errors is a very methodical process: list the measurements you made, then try, based on your knowledge of the instruments and methods used to make them, to determine the limits on the accuracy of each measurement. If you measured a length of an object with a meter stick, for example, how fine were the gradations on the meter stick? How much additional error might have been introduced by trying to line the object up with the markings on the meter stick by eye? If you measured a time with a stopwatch, how might your reaction time have contributed to the error in the time measurement? For these various sources of error, try to estimate by how much and in what direction the corresponding measurement may have been thrown off.

Was the error random or systematic?[10] Was friction likely to have affected your results, and if so, was there any evidence — such as speeds consistently lower than expected — of such an effect in your data?

For many errors, you will be able to estimate the magnitude of the error quantitatively as well as ascertain qualitatively how the error would have affected your results. For example, if you measure an object's length by lining it up with an ordinary meter stick, you would probably be confident of your length measurement to at least the 1 mm gradations on the stick. You might even argue that you could line the object up carefully enough to determine the length, not just to ± 1 mm, but to ± 0.5 mm. Either way, this error in the length measurement would be random, and if used, for example, to calculate a volume, would therefore be as likely to make your result for the volume too high as too low.

Sometimes you will have no way of estimating the magnitude of an error quantitatively and will only be able to ascertain the qualitative effect of the error. In such cases, you should explicitly note that it is not possible to estimate the error's magnitude numerically, then discuss the error's effect qualitatively and note whether there is evidence of such an effect in your results.[11] It is, for example, often obvious that friction is present, and you know that friction always tends to slow things down, but you usually won't be able to determine the extent of retardation quantitatively.

Once you have estimates of the error in each measurement, you can propagate these errors by the methods of §1.7 and derive an estimate of the corresponding overall error in your results, just as we did in the density example above.

1.6 Statistical Estimation of Error

Whenever possible, you should use the methods of the previous section to arrive at an informed quantitative estimate the magnitude of the error in each of your measurements. This is, however, not always possible; as noted in the previous section, there may be some errors that you cannot assess quantitatively. But if the measurement in question can be meaningfully repeated, it is still possible to obtain a quantitative estimate of the error by doing a statistical analysis of the results of these repeated measurements.

Suppose, for example, you are timing the period of a pendulum with a digital stopwatch. Some of the error in your time measurement will be due to human reaction time and judging when the pendulum is at the very end of a swing. You can minimize this error by measuring the total time for multiple swings of the pendulum and then dividing this total time by the number of

[10] To posit that an error was systematic, you must have a reason. You cannot, for example, know that a watch ran fast unless you compared it to a watch known to be calibrated correctly.

[11] This can be a bit tricky and will be discussed further in §1.10 and §1.11. You have to be very careful not to draw unjustified conclusions.

swings to obtain a result for the period: since the human timing error occurs only twice, at the beginning and end of the interval being timed, this error will become *proportionally* smaller for longer intervals. But the error is still present, and you might find it difficult to estimate its magnitude quantitatively by the methods of the previous section. There may also be error due to the limited accuracy of the timing mechanism inside the stopwatch. Even if we assume that the stopwatch is calibrated correctly, so that there is no systematic error, just because it displays times to $\frac{1}{100}$ sec doesn't necessarily mean its timing is always accurate to $\frac{1}{100}$ sec; unless you have some knowledge of the stopwatch's inner workings, you have no way of knowing how accurate it is.[12]

You can, however, sensibly repeat your time measurement: because of the above random errors, there will likely be some variation in your measured times, and, as we will see in §1.8, it is possible to derive a quantitative estimate of the error in the time measurement by analyzing the spread in your time values statistically. The relevant statistical quantity turns out to be what is known as the standard error.

Suppose, for example, that you have no life and make six measurements of the time for 100 periods of the pendulum, obtaining, in seconds,

204.11	197.23	199.10	200.05	202.45	199.87

Dividing by 100, the corresponding results for the time for one period are, again in seconds,[13, 14]

2.0411	1.9723	1.9910	2.0005	2.0245	1.9987

As your result for the period, you would of course want to use the mean (that is, the average) of these results:[15]

$$\overline{T} = \frac{2.0411 + 1.9723 + 1.9910 + 2.0005 + 2.0245 + 1.9987}{6} = 2.00468\,\text{sec}$$

As the estimate of the error, you would use the standard error of these results, which can be calculated by the method explained in §1.8. You have not, of course, gotten to §1.8 yet, but when you do you will see that you first determine the average $\overline{T^2}$ of the squares of your results,

$$\overline{T^2} = \frac{2.0411^2 + 1.9723^2 + 1.9910^2 + 2.0005^2 + 2.0245^2 + 1.9987^2}{6} = 4.01926\,\text{sec}^2$$

[12] Electronic stopwatches are in fact accurate to far better than the $\frac{1}{100}$ sec resolution they may display, but the point still stands in principle.

[13] The number of periods is an exact number; it has no error.

[14] If you used a calculator to perform this division, please do bring this to our attention so that we can beat you to death with it.

[15] There are various notations for averages; we will consistently use a bar over a quantity to denote its average value. And we have kept an extra significant figure on the result for the mean time as an intermediate result.

The standard deviation σ_{n-1} and standard error σ_μ of your $n = 6$ time measurements are then, according to eqq. (10) and (12) of §1.8,[16]

$$\sigma_{n-1} = \sqrt{\frac{n}{n-1}\left(\overline{T^2} - \overline{T}^2\right)} = \sqrt{\frac{6}{5}\left(4.01926 - 2.00468^2\right)} = 0.02493\,\text{sec}$$

$$\sigma_\mu = \frac{\sigma}{\sqrt{n}} = \frac{0.02493}{\sqrt{6}} = 0.01018\,\text{sec}$$

You would therefore quote your result for the period, including the standard error as the estimated error, as

$$2.00468 \pm 0.01018\,\text{sec}$$

if you were regarding it as an intermediate result to be used in subsequent calculations, or as

$$2.0047 \pm 0.0102\,\text{sec}$$

if you were regarding it as a final result and thus rounding back to the five significant figures in the original data.[17]

Bear in mind, however, that doing a blind statistical analysis of repeated measurements and using the standard error as an estimate of the error in those measurements is not as good as deriving an error estimate informed by knowledge of the instrument and of the method used to make the measurement, and indeed such statistical analyses are not even possible unless the measurement in question can be meaningfully repeated. If, for example, you are measuring the length of an object with a meter stick, you line the meter stick up with the object as carefully as you possibly can and record this single measurement for the length — unless you are performing the measurement during an earthquake, or are coming off a roaring bender and have a bad case of the shakes, there would have to be something seriously wrong with you to expect to get a different result for the length by repeating the measurement. Analysis of the error in such a measurement therefore cannot be carried out statistically.

1.7 Error Propagation

The object of the game is to determine the error in a result derived from one or more measurements.

[16] Here we have, as we will discuss further when we arrive at the next footnote, gotten a little funky with significant figures: the subtraction would in fact kill most of the significant figures in σ_{n-1} and σ_μ. But significant figures are, as it turns out, not all they're cracked up to be. And at any rate the reality is that everyone calculates everything to however many figures his or her calculator, computer, or slide rule gives and worries about significant figures only when it comes time to quote results.

[17] Note that in both cases we have matched the number of figures in the error to the number of figures in the time value: it would make no sense to quote, for example, 2.0047 ± 0.01018, with an error extending out to the hundred-thousandth place, when the hundred-thousandth figure in 2.0047 is already totally uncertain. And it would, similarly, make equally little sense to quote 2.0047 ± 0.010.

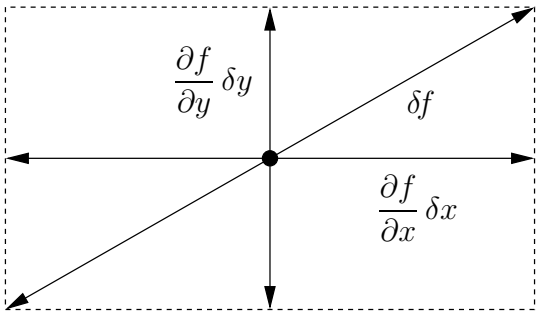

Figure 1: Combining Errors in Quadrature

We will use the notation δ for absolute errors and ϵ for relative errors. Thus for a measurement

$$x = 1.25 \pm 0.05\,\text{cm}$$

the relative error is

$$\frac{0.05}{1.25} = 0.04$$

and we will write

$$x = 1.25\,\text{cm} \qquad \delta x = 0.05\,\text{cm} \qquad \epsilon_x = \frac{\delta x}{x} = 0.04$$

Suppose that two measurements, x and y, are used to calculate a result that is a function $f(x, y)$ of x and y and that these measurements x and y have absolute errors δx and δy, respectively. We want to determine the error δf in the result $f(x, y)$ due to the errors in x and y.

If the errors δx and δy are small enough, we could use the simple linear approximation

$$\delta f \approx \frac{\partial f}{\partial x}\,\delta x + \frac{\partial f}{\partial y}\,\delta y$$

If x and y are independent, it is, however, conventional to treat them like orthogonal axes, as shown in fig. (1), and to use the Pythagorean theorem to obtain the error δf, so that

$$\delta f = \sqrt{\left(\frac{\partial f}{\partial x}\,\delta x\right)^2 + \left(\frac{\partial f}{\partial y}\,\delta y\right)^2} = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 \delta x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \delta y^2} \qquad (1)$$

This may seem slightly hokey, and you could come up with other schemes, but the idea is that independent measurements and their associated errors are in a sense like Cartesian axes: when the variables x and y are independent, algebraic shifts in the value of x have no effect on the value of y, just as geometric shifts along the x axis have no effect on your y coordinate, and vice versa. It is as though you could stray horizontally by a maximal "distance"

$$\frac{\partial f}{\partial x}\,\delta x$$

and vertically by

$$\frac{\partial f}{\partial y}\,\delta y$$

which would always leave you somewhere inside the dashed box of fig. (1). The maximal "distance" you could therefore stray from the center is the length of the half-diagonal. Errors combined in this way are said to be combined in quadrature.

For some of the more commonly occurring functional relationships, eq. (1), when you work out the derivatives, yields

$$\delta f = \left|n x^{n-1}\right|\,\delta x \qquad (f = x^n) \tag{2a}$$

$$\delta f = \sqrt{\delta x^2 + \delta y^2} \qquad (f = x \pm y) \tag{2b}$$

$$\delta f = \sqrt{y^2\,\delta x^2 + x^2\,\delta y^2} \qquad (f = xy) \tag{2c}$$

$$\delta f = \sqrt{\left(\frac{\delta x}{y}\right)^2 + \left(\frac{x\,\delta y}{y^2}\right)^2} \qquad \left(f = \frac{x}{y}\right) \tag{2d}$$

The error-propagation relations (2a), (2c), and (2d) can be more elegantly expressed in terms of the relative errors

$$\epsilon_f = \frac{\delta f}{f} \qquad \epsilon_x = \frac{\delta x}{x} \qquad \epsilon_y = \frac{\delta y}{y}$$

If, for example, we divide eq. (2c) through by $f = xy$, we obtain

$$\begin{aligned}
\epsilon_f = \frac{\delta f}{f} &= \frac{\sqrt{y^2\,\delta x^2 + x^2\,\delta y^2}}{xy} \\
&= \sqrt{\frac{y^2\,\delta x^2 + x^2\,\delta y^2}{x^2 y^2}} \\
&= \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2} \\
&= \sqrt{\epsilon_x^2 + \epsilon_y^2}
\end{aligned}$$

Doing likewise for eqq. (2a) and (2d), we arrive at

$$\epsilon_f = |n|\,\epsilon_x \qquad (f = x^n) \tag{3a}$$

$$\epsilon_f = \sqrt{\epsilon_x^2 + \epsilon_y^2} \qquad \left(f = xy \text{ or } \frac{x}{y}\right) \tag{3b}$$

As you can see from eqq. (2) and (3), when a variable is raised to a power, the relative error is multiplied by the absolute value of that power, and when

two variables are combined multiplicatively — that is, by either multiplication or division — , their relative errors are combined in quadrature. We can thus easily extend these rules to more complicated functional relationships or to functions of more than one variable. In the relation for the volume V of a rectangular box of length ℓ, width w, and height h, for example, the variables ℓ, w, and h are combined multiplicatively: $V = \ell w h$. To get the relative error ϵ_V in the volume we therefore simply combine the relative errors in ℓ, w, and h in quadrature:

$$\epsilon_V = \sqrt{\epsilon_\ell^2 + \epsilon_w^2 + \epsilon_h^2}$$

And in the pendulum relationship

$$g = \frac{4\pi^2 \ell}{T^2}$$

we have ℓ/T^2. By eq. (3a), the relative error in T^2 is twice the relative error in T: $\epsilon_{T^2} = 2\epsilon_T$. Since T^2 is then combined multiplicatively with ℓ, overall we have

$$\epsilon_g = \sqrt{\epsilon_\ell^2 + (2\epsilon_T)^2} \tag{4}$$

Rules for propagating errors for relations that cannot be handled by eqq. (2) and (3), such as those involving transcendental functions, can always be derived by applying eq. (1).

As a numerical example, suppose we have measured the sides of a rectangular box to an accuracy of 1 mm = 0.001 m and found that

$$\begin{aligned} \text{length } \ell &= 1.200 \pm 0.001 \text{ m} \\ \text{width } w &= 1.234 \pm 0.001 \text{ m} \\ \text{height } h &= 0.555 \pm 0.001 \text{ m} \end{aligned}$$

With three significant figures as a final result, the volume V is

$$V = 1.200 \times 1.234 \times 0.555 = 0.822 \text{ m}^3$$

The absolute and relative errors in the length, width, and height measurements are

$$\begin{aligned} \delta_\ell &= 0.001 \text{ m} & \epsilon_\ell &= \frac{\delta_\ell}{\ell} = \frac{0.001}{1.200} = 0.083\% \\ \delta_w &= 0.001 \text{ m} & \epsilon_w &= \frac{\delta_w}{w} = \frac{0.001}{1.234} = 0.081\% \\ \delta_h &= 0.001 \text{ m} & \epsilon_h &= \frac{\delta_h}{h} = \frac{0.001}{0.555} = 0.18\% \end{aligned}$$

where we have, since they are intermediate results, kept one extra figure on the relative errors. The relative and absolute errors in the result for the volume are therefore

$$\begin{aligned} \epsilon_V &= \sqrt{\epsilon_\ell^2 + \epsilon_w^2 + \epsilon_h^2} = \sqrt{0.083^2 + 0.081^2 + 0.18^2} = 0.21\% \\ \delta_V &= V \epsilon_V = 0.822 \times 0.21\% = 0.002 \text{ m}^3 \end{aligned}$$

and we would quote our result for the volume as $0.822 \pm 0.002 \text{ m}^3$.

1.8 Mean, Standard Deviation, Standard Error, & Confidence Levels

Suppose you have a set $\{x_i\}$, $i = 1, 2, \ldots, n$, of values obtained by repeating some particular measurement x. The mean μ of these n measurements is just their arithmetic average,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5}$$

To estimate the error in the measurement x, we want somehow to quantify the variation in the values of the x_i: the spread in these values will give us an indication of the uncertainty and hence inaccuracy of the measurement. One means of doing this is the standard deviation, which is defined as root-mean-square difference of the values $\{x_i\}$ from their mean and which we will denote by σ_n:[18]

$$\sigma_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{6}$$

In eq. (6), we take the deviation $x_i - \mu$ of each measurement from the mean, square this deviation ($(x_i - \mu)^2$), average these squared deviations ($\frac{1}{n}\sum(x_i - \mu)^2$), and then take the square root: hence the term *root-mean-square*. The squared deviations $(x_i - \mu)^2$ are averaged rather than the deviations $x_i - \mu$ themselves because we want a measure of how far x_i is from the mean μ without regard to whether that deviation is in the positive or negative direction.[19]

It is common to denote the average of a quantity by putting a bar over it. Thus the mean μ of the values $\{x_i\}$ is sometimes also written $\overline{x}$. And $\overline{x^2}$ would stand for the average value of the quantity x^2 — that is, what you get when you square all the $\{x_i\}$ and then average these squares.[20]

Eq. (6) is the intuitive definition of the standard deviation, but there is another relation for the standard deviation that is much more convenient for

[18] The reason for the subscript n will become clear shortly.

And while we're at it, note that the standard deviation, while it indicates the spread in the data values, is *not* the average distance of the data points from the mean. This is demonstrated in Appendix C and is the reason why those trying to be precise may speak of standard deviation as the "typical" distance of the data points from the mean or engage in similar verbal gymnastics.

[19] We could instead average the absolute values of the deviations $x_i - \mu$, but the square is conventionally used rather than the absolute value because, in addition to other compelling reasons considerably beyond the scope of what we are doing, absolute value is not a nice function mathematically. And of course averaging the deviations $x_i - \mu$ directly would just be silly: you would always get zero, because the data points $\{x_i\}$ are right at the mean value *on average*:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu) = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n}\mu = \mu - \frac{1}{n}(n\mu) = 0$$

[20] As opposed to $\overline{x}^2$, which stands for squaring the mean value $\overline{x}$ of the $\{x_i\}$.

calculation: if we use bar notation and expand the square in eq. (6), we obtain

$$\begin{aligned}\sigma_n &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \\ &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2\overline{x}x_i + \overline{x}^2)} \\ &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\frac{1}{n}\sum_{i=1}^{n}\overline{x}x_i + \frac{1}{n}\sum_{i=1}^{n}\overline{x}^2}\end{aligned}$$

Since the mean $\overline{x}$ is just a constant, it can be pulled out of the summations, so that we have

$$\begin{aligned}\sigma_n &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\overline{x}\frac{1}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}\overline{x}^2\sum_{i=1}^{n}1} \\ &= \sqrt{\overline{x^2} - 2\overline{x}\,\overline{x} + \frac{1}{n}\overline{x}^2\,n} \\ &= \sqrt{\overline{x^2} - \overline{x}^2} \qquad (7)\end{aligned}$$

Eq. (7) makes calculation of the standard deviation much easier: you just take the average of the x_i^2, substract the square of the mean, and take the square root.

The standard deviation σ_n that we have been discussing is known more precisely as the standard deviation with n weighting, that is, the standard deviation you obtain when you take the root-mean-square deviation by averaging over *all* n *data points*, corresponding to the division by n in eq. (6) and hence to the n subscript on σ_n. The values of the mean and standard deviation that you calculate from your data points are, however, only an approximation to the true values of the mean and standard deviation, which are the values that would be obtained in the limit of infinitely many data points. To distinguish them from the true mean and standard deviation, the mean and standard deviation you calculate from your data points are often called the sample mean and sample standard deviation.

The n data points themselves, being independent of each other, are sometimes referred to as degrees of freedom. When you use those n data points to calculate the value of the sample mean, you still have n data points, but their number of degrees of freedom is reduced to $n-1$ because they are no longer fully independent of each other: there is now the relation (5) among them. In effect, it is as though one of the n degrees of freedom has been eaten up by the sample mean. It is therefore customary when calculating the sample standard deviation to average, not over all n data points, but over the remaining $n-1$ degrees of freedom, so that eq. (6) becomes instead

$$\sigma_{n-1} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2} \qquad (8)$$

This new animal σ_{n-1} is known as the standard deviation with $n-1$ weighting and is, as you can see by comparing eqq. (6) and (8), related to the standard deviation σ_n with n weighting by

$$\sigma_{n-1} = \sigma_n \sqrt{\frac{n}{n-1}} \tag{9}$$

Thus eq. (7) becomes

$$\sigma_{n-1} = \sqrt{\frac{n}{n-1}\left(\overline{x^2} - \overline{x}^2\right)} \tag{10}$$

This is, however, still not the end of the story: there is yet another kind of standard deviation, known as the standard deviation of the mean or, in the context of error analysis, the standard error. You would expect that the more data points you have, the closer your sample mean will be to the true mean. Although the root-mean-square deviation of your data points will not change much as you collect more of them, your result for the *average* will get progressively closer to the true average. The proof is considerably beyond what we can or want to get into here,[21] but it can be shown that if you were to repeat your set of measurements infinitely many times, so that you had infinitely many sets of data, each set consisting of n data points, and you then calculated the sample mean of each of these sets, then those sample means would have a standard deviation σ_μ given by

$$\sigma_\mu = \frac{\sigma}{\sqrt{n}} \tag{11}$$

where σ is the standard deviation of the data points. Since we use the mean value $\overline{x}$ as our result for the measurement x when we are analyzing errors statistically, our concern is not with how far our individual data points typically fall from this mean, but how far our mean $\overline{x}$ falls from the true mean. We will therefore use the standard error σ_μ of eq. (11) as our estimate of the absolute error δx in our measurement x, specifically

$$\sigma_\mu = \frac{\sigma_{n-1}}{\sqrt{n}} \tag{12}$$

It is important to recognize, however, that what the standard error converges to in the limit of a large number of measurements is the true mean, which is not necessarily the same as the true value of the quantity being measured: the standard error converges to the true mean because of the tendency of random errors to cancel each other on average, but no such cancellation occurs for systematic errors; if there is any systematic error in the instrumentation or method used to make a measurement, that systematic error will persist even when the results of numerous measurements are averaged. There is no way to take systematic error into account without knowledge of its nature

[21] For those interested, this proof is outlined in Appendix B.

# stderr	Chance	% CL	# stderr	Chance	% CL
0.5	0.617	38.3	3.0	2.7×10^{-3}	99.7
1.0	0.317	68.3	4.0	6.3×10^{-5}	99.99
1.5	0.134	86.6	5.0	5.7×10^{-7}	99.9999
2.0	0.046	95.4	6.0	2.0×10^{-9}	99.9999998
2.5	0.012	98.8	7.0	2.6×10^{-12}	99.999999997

% CL	68	90	95	99
# stderr	1.00000	1.64485	1.95996	2.57583

Table 1: Confidence Levels

and magnitude: we can correct for a systematic error only when we know in what direction and by how much it affects the measurement.

Although in different ways, all three of σ_n, σ_{n-1}, and σ_μ are measures of the variation or spread in your data values. Data with completely random sources of error often fits very well a distribution called the normal distribution, also informally known, because of its shape when plotted, as the bell curve. We can't really get into why this should be so or even what exactly a normal distribution is,[22] but it turns out that if, as we will always assume, your data points are normally distributed — that is, if they fit a normal distribution — , there is a 68% chance that the true value of the quantity being measured will fall within one standard deviation of the mean. The farther you go out on either side of the mean, the more likely that the true value will fall within this wider range. The chance of the true value falling within 1.64485σ of the mean is 90%; within 1.95996σ, 95%; and within 2.57583σ, 99%. In other words, if you use 2.57583 standard errors as your error estimate, there is a 99% chance that the true value of the measured quantity falls within this error range, and you are said to have a 99% confidence level.[23]

Table (1) shows probabilities and confidence levels for given ranges on either side of the mean: the columns *# stderr*, *Chance*, and *% CL* indicate the range to either side of the mean, the probability of the true mean falling outside of this range (as a fraction of 1, not as a percentage), and the corresponding confidence level (which is just 1 minus the entry in the previous column, expressed as a percentage).

[22] Again, those interested should see Appendix B.

[23] The chance that the true value of g falls outside of a given number of standard errors of your result is determined by the cumulative distribution function of the normal distribution and its inverse function, which involve integrations that have to be carried out numerically and are beyond our scope.

For a given set of data points, there is a unique mean and standard error, and you will always use the mean as the value of the measurement. You can, however, express the error in the measurement at various confidence levels. Recall, for example, that on p.9 the average of the time measurements in our pendulum example was was 2.0047 sec and their standard error 0.0102 sec (0.01018 before rounding). If you chose to work at the 68% confidence level, you would use a single standard error as the error in the time measurement and would quote your result as

$$2.0047 \pm 0.0102\,\text{sec} \tag{13}$$

meaning that there is a 68% chance that the true value of the time falls between 1.9945 and 2.0149 sec.[24] If instead you chose to work at the 99% confidence level, you would use 2.57583 standard errors as the error in the time measurement:

$$2.57583 \times 0.01018 = 0.0262\,\text{sec}$$

You would then quote your result as

$$2.0047 \pm 0.0262\,\text{sec} \tag{14}$$

meaning that there is a 99% chance that the true value of the time falls between 1.9785 and 2.0309 sec.

It is very important to understand to that both (13) and (14) represent the same measurement *made to the same accuracy*; (14) is *not* less accurate than (13). It is just that this error is expressed with different confidence levels and therefore differs in its interpretation. To be 68% confident that the true value of the time falls within the range 2.0047 ± 0.0102 sec is the same as being 99% confident that the true value of the time falls within the range 2.0047 ± 0.0262 sec: that the range in (14) is wider than that in (13) corresponds, not to (14) being less accurate than (13), but to the greater confidence that the true value of the time falls within the specified range.

1.9 A More Complete Example of Error Analysis

Now that we have covered all the various technical aspects of error analysis, as a more complete example of performing such an analysis and interpreting its results, consider the following fictional experiment:

You are really bored one day and decide to while away the time by using a pendulum to obtain a result for the acceleration g due to gravity. As noted earlier, the relation for g in terms of the pendulum's period T and length ℓ is

$$g = \frac{4\pi^2\ell}{T^2}$$

[24] Which of course also means that there is still a 32% chance that the true time falls outside of the range 2.0047 ± 0.0102 sec, either less than 1.9945 sec or more than 2.0149 sec.

You set up your pendulum and use a tape-measure to measure its length ℓ, which you find to be 0.995 m. By carefully taking into account the fineness of the gradations on the tape, the difficulty of holding the tape straight along the pendulum, and the difficulty of aligning the ends of the pendulum string with the markings on the tape, you arrive at a careful estimate of ± 3 mm error in your length measurement. Your result for the length of the pendulum is thus

$$\ell = 0.995 \pm 0.003\,\mathrm{m}$$

where $\delta\ell = 0.003$ m is the absolute error in your length measurement. The corresponding relative error ϵ_ℓ is

$$\epsilon_\ell = \frac{\delta\ell}{\ell} = \frac{0.003}{0.995} = 0.0030$$

where we have kept an extra significant figure on ϵ_ℓ because it is an intermediate result that we will later be using to propagate this error.

As discussed in §1.6, the random error in the timing of the pendulum's period has to be dealt with statistically, but it will be proportionally smaller the larger the number of swings you time. You therefore make repeated measurements of the total time for 100 periods — six such measurements, before the tedium saps you of your will to live. To save ourselves the trouble of having to calculate the mean and standard error, we will suppose that these six total times come out the same as those used in the example on p.8: in seconds,

204.11	197.23	199.10	200.05	202.45	199.87

so that the corresponding results for the period T are, again in seconds,

2.0411	1.9723	1.9910	2.0005	2.0245	1.9987

In §1.6 we found, by using eqq. (5), (10), and (12), that the mean $\overline{T}$ and standard error σ_μ of these results for the period are

$$\overline{T} = 2.00468\,\mathrm{sec} \qquad\qquad \sigma_\mu = 0.01018\,\mathrm{sec}$$

where we have kept one more figure in $\overline{T}$ as an intermediate result. Note that the standard error has been rounded off so that its last significant figure (the hundred-thousandth place) aligns with the last significant figure in the mean period $\overline{T}$: digits in the standard error that go beyond the figures in the mean are meaningless.

What you use for your estimate of the error in the period T will depend on how conservative your estimate of the error in the length ℓ was: If your ± 3 mm was merely the *likely* error in ℓ, for consistency you probably want to use something like one standard error for your estimate of the error in T, since one standard error corresponds to a confidence level of 68%, and 68% would represent a fair likelihood as betting odds go. If, on the other hand,

your ± 3 mm was a *worst-case* estimate of the error in ℓ, for consistency you probably want to use something like 2.57583 standard errors for your estimate of the error in T, since 2.57583 standard errors correspond to a confidence level of 99%, and 99% is a virtual cinch as betting odds go.[25]

If your estimates are of likely error, you would therefore use a single standard error as the absolute error in $\overline{T}$, so that your result for $\overline{T}$ and its absolute and relative errors are[26]

$$\overline{T} = 2.00468\,\text{sec} \qquad \delta\overline{T} = 0.01018\,\text{sec} \qquad \epsilon_{\overline{T}} = \frac{0.01018}{2.00468} = 0.005078$$

If instead your estimates were of worst-case error, you would use 2.57583 standard errors as the absolute error in $\overline{T}$, so that your result for $\overline{T}$ and its absolute and relative errors would be

$$\overline{T} = 2.00468\,\text{sec} \qquad \delta\overline{T} = 2.57583 \times 0.01018 = 0.02622\,\text{sec}$$
$$\epsilon_{\overline{T}} = \frac{0.02622}{2.00468} = 0.01308$$

In either case you would, having the same result for $\overline{T}$, of course get the same result for g:

$$g = \frac{4\pi^2\ell}{T^2} = \frac{4\pi^2 \times 0.995}{2.00468^2} = 9.774\cdots = 9.77\,\text{m/s}^2$$

where we have noted that, since in the original length data had three significant figures and our time measurements five significant figures, there should be three significant figures in our final result for g.

The error-propagation relation for our result for g, as discussed on p.12, is given by eq. (4):

$$\epsilon_g = \sqrt{\epsilon_\ell^2 + (2\epsilon_T)^2}$$

If your estimate is of likely error, this will work out to

$$\epsilon_g = \sqrt{(0.0030)^2 + (2 \times 0.005078)^2} = 0.01059$$

so that the absolute error in g will be

$$\delta g = \epsilon_g g = 9.774 \times 0.01059 = 0.10\,\text{m/s}^2$$

and you would therefore quote your result for g as

$$g = 9.77 \pm 0.10\,\text{m/s}^2 \tag{15}$$

[25] *Likely* and *worst-case* error are *not* conventional technical terms, nor is the equating of these errors to the 68% and 99% confidence levels conventional; they are merely terms and prescriptions we will use in this course.

[26] Here we have again, for the reasons discussed in footnote 17 on p.9 and its predecessor, gotten a bit funky with the significant figures in the relative error. But this is rectified when we calculate the absolute error in g. No blood, no foul.

Again note that the error has been rounded off so that its last significant figure (the hundredth place) aligns with the last significant figure in the result for g.

You would interpret your result $g = 9.77 \pm 0.10\,\mathrm{m/s^2}$ as meaning that although you arrived at $9.77\,\mathrm{m/s^2}$ for g, there is about a 68% likelihood that the true value of g lies somewhere between $9.67\,\mathrm{m/s^2}$ and $9.87\,\mathrm{m/s^2}$. Your result would therefore certainly be regarded as consistent with a reported value of $9.80\,\mathrm{m/s^2}$, as it would with reported values anywhere in the range from $9.67\,\mathrm{m/s^2}$ to $9.87\,\mathrm{m/s^2}$. But there is still a 32% chance that the true value of g lies outside this range. As you can see from Table (1) on p.16, the odds of the true value of g falling farther than one standard error from your result decrease rapidly as the discrepancy goes out to two, three, four, or more standard errors. There is, for example, only a $2 \times 10^{-9} = 0.0000002\%$ chance of the true value of g being six or more standard errors — that is, $6 \times 0.10 = 0.60\,\mathrm{m/s^2}$ — away from $9.77\,\mathrm{m/s^2}$. Your result would therefore certainly be regarded as inconsistent with a reported value of $9.17\,\mathrm{m/s^2}$. But reported values that fall only slightly outside the range for your 68% confidence level might well still be consistent with your result, so that you could not, for example, definitively conclude that a reported value of $9.60\,\mathrm{m/s^2}$ was inconsistent with your result.

If instead your estimate were of worst-case error, ϵ_g would work out to

$$\epsilon_g = \sqrt{(0.0030)^2 + (2 \times 0.01308)^2} = 0.02633$$

so that the absolute error in g will be

$$\delta g = \epsilon_g g = 9.774 \times 0.02633 = 0.26\,\mathrm{m/s^2}$$

and you would therefore quote your result for g as

$$g = 9.77 \pm 0.26\,\mathrm{m/s^2} \tag{16}$$

In this case you would interpret your result $g = 9.77 \pm 0.26\,\mathrm{m/s^2}$ as meaning that you arrived at $9.77\,\mathrm{m/s^2}$ for g, with about a 99% likelihood that the true value of g lies somewhere between $9.51\,\mathrm{m/s^2}$ and $10.03\,\mathrm{m/s^2}$. There is only a 1% chance that the true value of g falls outside this range. To the extent that a 1% chance is regarded as negligibly small, your result would therefore be regarded as inconsistent with reported values smaller than $9.51\,\mathrm{m/s^2}$ or greater than $10.03\,\mathrm{m/s^2}$.

To repeat a critically important point: both (15) and (16) represent the same result *made to the same accuracy*; (16) is *not* less accurate than (15). It is just that the error is expressed with different confidence levels and therefore differs in its interpretation. To be 68% confident that the true value of g falls within the range $g = 9.77 \pm 0.10\,\mathrm{m/s^2}$ is the same as being 99% confident that the true value of g falls within the range $g = 9.77 \pm 0.26\,\mathrm{m/s^2}$: that the range in (16) is wider than that in (15) corresponds, not to (16) being less accurate than (15), but to the greater confidence that the true value of g falls within the specified range.

Although, for the sake of comparing their interpretations, we have worked out the error at both the 68 and 99% confidence levels, when you perform an error analysis you will do so for just *one* confidence level. The only restriction on your choice, as discussed in the above pendulum example, is the need for consistency between your statistically estimated and your other errors: it makes no sense to combine errors of differing confidence levels when you propagate errors.

1.10 The Many Guises of Satan

Experience has shown that as you gleefully skip along through your lab work there are many pitfalls into which you are likely fall. People your age are, of course, notoriously averse to believing that bad things can happen to you, but the wise will carefully heed the following hints and warnings.

1. When quoting the values of measurements or results, you should always specify not only the value and its units, but also the absolute error and confidence level. The result for a velocity, for example, would be quoted in the form

   $$X.XXX \pm Y.YYY \text{ m/s at the } ZZ\% \text{ confidence level}$$

   or the like.

2. (a) Significant figures prevent, not cause, rounding error. Preventing rounding error is in fact the whole point of the rules for significant figures.

   (b) Remember that for this purpose you should keep an extra figure in intermediate results.

   (c) And as we've noted in the preceding sections, you should be sure that the significant figures in your error are matched to those of the value: 1.234 ± 0.056, for example, and not 1.234 ± 0.0567 or 1.234 ± 0.05.

3. (a) Consider all and only those errors that are likely both to have occurred and to have significantly affected your measurements. Conversely, do not cite fantastic sources of error.

   We should not, for example, be reading things in your lab report like "the computer might have malfunctioned" or "the relative humidity in China was high when we measured the distance." Unless of course you performed the experiment in China and you used meter sticks made of balsa wood.

   (b) A warped meter stick will cause error, but you should never say simply that the meter stick *might have been* warped: you could have definitively ascertained just by looking lengthwise down the stick whether it was significantly warped or not. And if you find that the meter stick you are using is warped, you should, if possible, find one that isn't warped and use that. Similarly, if the end of your meter stick has been chewed

off, you should measure from where the scale on the stick becomes good rather than from its end.

4. While you can and should consider errors of human origin, such as the error in trying to align a meter stick by eye, never cite simply "human error"; it's too vague to be meaningful.[27]

5. (a) In your report, you should give a full, detailed explanation of how you made each measurement.

   You should not, for example, simply quote that your result for a length measurement was 1.23 m; you need to explain how you made the length measurement — what instrument you used, and between what two points you measured.

   (b) i. When it is possible to estimate the magnitude of an error quantitatively, in your report you should also give a full, detailed explanation of how you have arrived at your error estimate.

   You should not, for example, simply quote that your error estimate for a length measurement was ± 0.01 m; you need to explain all the considerations that led to your estimate of ± 0.01 m error.

   ii. When you do not have enough information to make a quantitative estimate of an error in a measurement, do not make up arbitrary figures for the error's magnitude.

   You should, however, try to reason out whether the error is random or systematic, and, if the latter, to further reason out the direction in which it would have affected your result, that is, whether it would have tended to increase or decrease the value of your result.

6. An experimental error is the limit on the accuracy of a measurement or result. A mistake is not an experimental error. If you made a mistake solving a quadratic equation, you should correct it. If you think you *might* have made a mistake solving a quadratic equation, you should check your calculation. And if you discover after the fact, when it is too late to correct it, that you made a mistake in the lab, you should discuss the mistake and how it might have affected your results, but be careful to distinguish it from experimental error.

7. (a) Sometimes there will be other values reported for an experimental result you have obtained. If you have obtained a value for the acceleration g due to gravity, for example, other lab groups will have obtained their own results for g, and there may also be a generally reported local value for g. In such cases, you can and should compare your result to these other results, and the issue is whether your result is or is not consistent with those results, as discussed at the end of the previous section.

[27] Included in this type of error are all the variations on the theme "my lab partner is a loser." Even if this is demonstrably true in a scientifically objective sense, you still need to be more specific.

i. Unless you did something wrong, your result must, to within your range of error, be correct. If your result is inconsistent with someone else's, then as far as you are concerned, it's the other guy that's wrong. The inconsistency might cause you to review your lab work, calculations, assumptions, etc., but unless you find — or someone else points out — a mistake in your work, you should not be concluding that your result is incorrect.

ii. Experimental error is *not* the difference between your result and someone else's. Anyone suggesting this will be dragged outside and shot.

    Similarly, to think that the validity or accuracy of your result is determined by how close you came to some generally reported value betrays a capital lack of understanding of experimental error.

    If you want to refer to the difference between your result and someone else's, you should use *discrepancy* or some other term that makes it clear you are not considering this difference to be an experimental error.

iii. There is no such thing as a "theoretical" value for g or for any other experimentally determined quantity.[28]

8. Be very careful of drawing unjustified conclusions. To conclude that an effect on your results was significant, you must know both that the effect occurred and that it was of significant magnitude.

   You will, for example, frequently know that there were frictional forces for which you did not account, but you will seldom have made any measurements to ascertain the magnitude of the effect of friction. If friction would have tended to lower your result, that your result came out lower than expected does not justify the conclusion that friction was *the* cause: you do not know whether the effect of friction was of significant magnitude, and even if it was, you must still rule out that your result came out low because of other effects or even simply because of random error. Nor would having your result come out high justify the conclusion that friction was of negligible magnitude: friction might in fact have significantly lowered your result, but with this lowering countered by other effects or masked by random error.[29]

   By all means you should note in your report that there was at least some degree of friction, and, if you can, further reason out qualitatively the direction in

[28] At least at the level at which we are working. Sometimes theory does make absolute predictions of physical quantities, such as the prediction of the value of the gyromagnetic ratio of the electron by quantum electrodynamics.

[29] Beware: Experience has shown that the temptation to conclude "error X would have made our result smaller, and our result was smaller than the expected value, therefore error X was responsible for our result coming out small" is almost beyond the power of human will to resist. Just further evidence that humans are separated from other animals principally by their capacity for irrationality.

which it would have affected your results, but you must be very careful about the conclusions you draw: whether friction definitely significantly affected your results, whether it might have significantly affected them, whether it might not have affected them significantly, or whether it definitely did not affect them significantly.

1.11 Exercises

1. Calculate the mean and standard error of

   9.9 9.5 10.1 10.5 10.2 9.8

   (You really shouldn't need a calculator to do this.)

2. While doing an experiment, you note that the ends of the meter stick you are using to make measurements are worn. There are no meter sticks in the lab without worn ends, and the nature of the measurement is such that you must measure from the end of the stick. You should

   (a) Break down in tears.

   (b) Complain bitterly to the administration about the abysmally poor condition of the physics labs.

   (c) Challenge a neighboring lab group to a sword fight with the worn meter stick.

   (d) Use a second meter stick to measure the amount of wear and adjust your length measurements accordingly.

   (e) Include this source of error in your error estimate, without adjusting your length measurements.

3. You find that your result for a physical quantity Q comes out lower than expected.

   (a) A warped meter stick would have thrown your length measurements off in such a way as to lower the value you obtain for Q.

      i. Are you justified in concluding that a warp in the meter stick *was* (at least partially) the cause of your obtaining a low value of Q?

      ii. Are you justified in concluding that a warp in the meter stick *might have been* (at least partially) the cause of your obtaining a low value of Q?

   (b) A warped meter stick would have thrown your length measurements off in such a way as to *raise* the value you obtain for Q and thus move it even farther from the generally reported value.

   Are you justified in concluding that the meter stick could not have been warped?

4. The generally reported value for the acceleration due to gravity at Gassy Swamp, New Jersey is $g = 9.802\,\mathrm{m/s^2}$. From your experiment, you obtain $g = 9.900 \pm 0.200\,\mathrm{m/s^2}$ at the 99% confidence level.

   (a) Does this mean that you are 99% sure that your result is consistent with the generally reported value?

   (b) Does this mean that there was some source of systematic error causing your result for g to come out high?

   (c) Does this mean that there was some random error causing your result for g to come out high?

# stderr	Chance	% CL	# stderr	Chance	% CL
0.5	0.617	38.3	3.0	2.7×10^{-3}	99.7
1.0	0.317	68.3	4.0	6.3×10^{-5}	99.99
1.5	0.134	86.6	5.0	5.7×10^{-7}	99.9999
2.0	0.046	95.4	6.0	2.0×10^{-9}	99.9999998
2.5	0.012	98.8	7.0	2.6×10^{-12}	99.999999997

Figure 1: Problem 5

We will suppose that you have obtained an experimental result for a physical quantity Q. The probability that the true value of Q falls a given number of standard errors away from your result for Q can be determined from what is known as the cumulative distribution function of the normal distribution. The above table shows the chance of the true value of Q differing from your result for Q by n or more standard errors for $n = 0.5, 1.0, 1.5, \ldots, 7.0$, along with the corresponding confidence level. Thus, for example, the chance of the true value of Q being 1.0 or more standard errors either above or below your result for Q would be $0.317 = 31.7\%$, corresponding to a $100 - 31.7 = 68\%$ confidence level.

5. The physical quantity Q is generally reported to have the value 2.00 in units of Zounds (Zd). Explain what your error and confidence level tell you about the accuracy of your result for Q, and indicate what you can conclude about your result's degree of consistency or inconsistency with the generally reported result 2.00 Zd, if your experimental measurements yield

   (a) $Q = 2.10 \pm 0.20$ Zd at the 68% confidence level.

   (b) $Q = 1.20 \pm 0.20$ Zd at the 99% confidence level.

   (c) $Q = 2.35 \pm 0.10$ Zd at the 68% confidence level.

   (d) $Q = 2.35 \pm 0.10$ Zd at the 99% confidence level.

   (e) $Q = 1.60 \pm 0.05$ Zd at the 68% confidence level.

   (f) $Q = 2.40 \pm 2.00$ Zd at the 68% confidence level.

   (g) $Q = 2.00 \pm 0.20$ Zd at the 68% confidence level.

   (h) $Q = 2.00 \pm 1.80$ Zd at the 68% confidence level.

   (i) $Q = 1.95 \pm 0.01$ Zd at the 68% confidence level.

   (j) $Q = 8.00 \pm 0.20$ Zd at the 68% confidence level.

   (k) $Q = 8.00 \pm 0.01$ Zd at the 68% confidence level.

   (l) $Q = 8.00 \pm 7.00$ Zd at the 68% confidence level.

# stderr	Chance	% CL	# stderr	Chance	% CL
0.5	0.617	38.3	3.0	2.7×10^{-3}	99.7
1.0	0.317	68.3	4.0	6.3×10^{-5}	99.99
1.5	0.134	86.6	5.0	5.7×10^{-7}	99.9999
2.0	0.046	95.4	6.0	2.0×10^{-9}	99.9999998
2.5	0.012	98.8	7.0	2.6×10^{-12}	99.999999997

Figure 1: Problem 6

6. In your experiment to determine the value of Q, you know that there was friction and that the effect of friction would have been to depress your result for Q — that is, that the value of Q you would have obtained in the absence of friction would have been higher than the value of Q yielded by your experiment. You have, however, no way of knowing the magnitude of the effect of friction. Determine whether you can conclude that

   i. friction was definitely at least partially the cause of the discrepancy between your result for Q and its generally reported value

   ii. friction was probably at least partially the cause of the discrepancy between your result for Q and its generally reported value

   iii. friction may have been at least partially the cause of the discrepancy between your result for Q and its generally reported value

   iv. friction probably did not significantly affect your result for Q

   v. friction definitely did not significantly affect your result for Q

   if you obtained, at the 68% confidence level,

   (a) $Q = 2.10 \pm 0.20$ Zd.

   (b) $Q = 1.20 \pm 0.20$ Zd.

   (c) $Q = 2.35 \pm 0.10$ Zd.

   (d) $Q = 2.00 \pm 0.20$ Zd.

   (e) $Q = 1.95 \pm 0.01$ Zd.

2 Fitting Curves

2.1 The Least-Squares Fit

Usually you have some idea what sort of functional relationship your data should obey, that is, what sort of curve your data points should fit: parabolic, cubic, exponential, confluent hypergeometric, etc. To fit your data to the curve, you need to determine the values of the curve's parameters that will, by some quantitative prescription, yield the closest fit. To fit a set of data points of the form (x, y) to the quadratic $y = ax^2 + bx + c$, for example, you need to determine the values of a, b, and c. While there are very sophisticated and efficient techniques for fitting sets of data points to general curves, these lie far enough beyond what we can or want to get into here that we will restrict ourselves to the special case that your data is expected to obey a linear relationship and your data points should therefore fit a straight line. But the technique for fitting straight lines, in addition to illustrating the basic considerations common to more sophisticated fits, will prove of more general application than you might at first think: even when the relationship is nonlinear, there is often a simple change of variables that will make the relationship linear. For free-fall from rest, for example, you expect the distance h fallen to be related to the time t of flight by $h = \frac{1}{2}gt^2$. Although the plot of h versus t should be parabolic, a plot of h versus t^2 should be a straight line. Restricting ourselves to linear plots will therefore not limit us as severely as it might seem.

Suppose then that we have a set of n data points (x_i, y_i), $i = 1, \ldots, n$, that we want to fit to the straight line $y = mx + b$. The object of the game is to determine the slope m and intercept b that will make the line pass as closely as possible to as many of the data points as possible. One way to accomplish this is to minimize the sum of the squared distances between the line and the data points — a method that is therefore known as the least-squares fit. At $x = x_i$, the value of the y coordinate of the data point (x_i, y_i) is y_i, while the value of the y coordinate on the line is $mx_i + b$, as shown in fig. (2). The squared

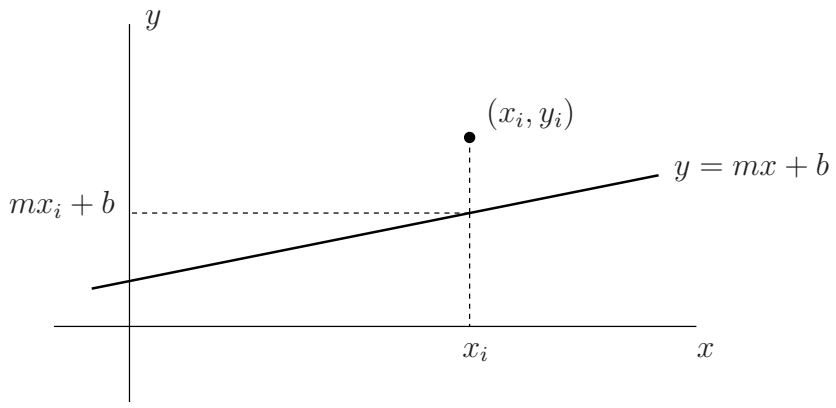

Figure 2: The Distance Between the Line and the Data Point

vertical distance $(\Delta y)_i$ between the line and data point (x_i, y_i) is thus [30]

$$(\Delta y)_i^2 = \Big(y_i - (mx_i + b)\Big)^2$$

and hence the sum of the squared vertical distances between the line and each of the n data points

$$\sum_{i=1}^{n} \Big(y_i - (mx_i + b)\Big)^2$$

We want to choose the slope m and intercept b that will minimize this sum, that is, that satisfy the maxima-minima relations

$$0 = \frac{\partial}{\partial m} \sum_{i=1}^{n} \Big(y_i - (mx_i + b)\Big)^2 \qquad 0 = \frac{\partial}{\partial b} \sum_{i=1}^{n} \Big(y_i - (mx_i + b)\Big)^2$$

Doing out these derivatives by the chain rule and doing a bit of gerrymandering yields

$$0 = \frac{\partial}{\partial m} \sum_{i=1}^{n} \Big(y_i - (mx_i + b)\Big)^2$$

[30] At this juncture, you may have two questions: why use the vertical distance, and why square it? For a straight line, it actually doesn't matter in which direction we measure the distances between the line and the data points. Fig. (3) shows two different directions — one vertical, the other tilted at some arbitrary angle — along which we might measure the distances between the data points and the best-fit line. Since the triangles formed in this way are similar,

$$\frac{a}{A} = \frac{b}{B}$$

and minimizing the sum of the squares of the "A" sides is therefore equivalent to minimizing the sum of the squares of the "B" sides. And we square the distances precisely because we want to deal with distances, not with displacements, which will be negative for the data points that lie under the line. For this purpose we could, of course, equally well use $|\Delta y|$ or $(\Delta y)^4$ or the like instead of $(\Delta y)^2$, but the use of the square is nicer, simpler, and conventional.

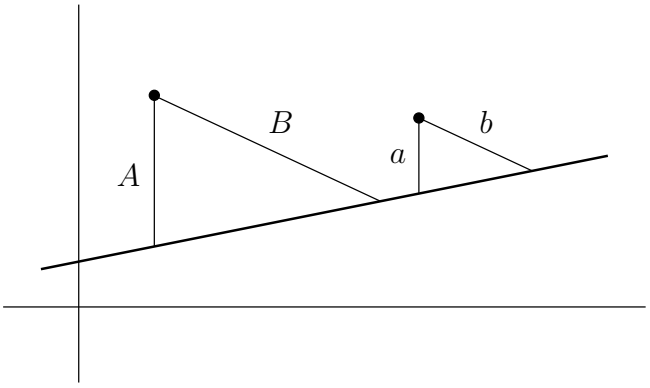

Figure 3: Direction Don't Matter None

$$= \sum_{i=1}^{n} 2\Big(y_i - (mx_i + b)\Big)\,\frac{\partial}{\partial m}(mx_i + b)$$

$$= \sum_{i=1}^{n} 2\Big(y_i - (mx_i + b)\Big)(-x_i)$$

$$= -2\sum_{i=1}^{n}(x_i y_i - mx_i^2 - bx_i)$$

$$= -2n\left(\frac{1}{n}\sum_{i=1}^{n}(x_i y_i - mx_i^2 - bx_i)\right)$$

$$= -2n\left(\overline{xy} - m\overline{x^2} - b\overline{x}\right)$$

and

$$0 = \frac{\partial}{\partial b}\sum_{i=1}^{n}\Big(y_i - (mx_i + b)\Big)^2$$

$$= \sum_{i=1}^{n} 2\Big(y_i - (mx_i + b)\Big)\,\frac{\partial}{\partial b}(mx_i + b)$$

$$= \sum_{i=1}^{n} 2\Big(y_i - (mx_i + b)\Big)(-1)$$

$$= -2\sum_{i=1}^{n}(y_i - mx_i - b)$$

$$= -2n\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - mx_i - b)\right)$$

$$= -2n\left(\overline{y} - m\overline{x} - b\right)$$

where as usual $\overline{x}$, $\overline{y}$, $\overline{xy}$, and $\overline{x^2}$ denote the average values of the quantities x, y, xy, x^2, respectively. Thus we arrive at the conditions

$$0 = \overline{xy} - m\overline{x^2} - b\overline{x} \qquad\qquad 0 = \overline{y} - m\overline{x} - b$$

Solving these simultaneous equations for m and b yields

$$m = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} \qquad\qquad b = \frac{\overline{y}\,\overline{x^2} - \overline{x}\,\overline{xy}}{\overline{x^2} - \overline{x}^2} \tag{17}$$

These are the values of the slope and intercept of the line that will most closely fit the data points.

As an example of a least-squares fit by hand to sample data, let us suppose we have a set of four data points: [31]

$$(1.00, 1.10) \qquad (2.00, 2.00) \qquad (3.00, 2.90) \qquad (4.00, 4.00)$$

[31] Fitting just four data points to a line is of course pathetic; in general, you want the number of data points to far exceed the number of parameters in the fit — in the case of a straight line, two (the slope and intercept). Each parameter in the fit in effect eats up one degree of freedom in the set of data points, and the fewer the degrees of freedom remaining after the fit, the less meaningful the fit. Doing an n parameter fit of n data points would be completely meaningless. This will be discussed further in the next section.

Just eyeballing these points, it would seem like the best fit should be something close to the line $y = x$. To apply eqq. (17), we will need the quantities

$$\overline{x} = \tfrac{1}{4}(1.00 + 2.00 + 3.00 + 4.00) = 2.500 \tag{18a}$$

$$\overline{y} = \tfrac{1}{4}(1.10 + 2.00 + 2.90 + 4.00) = 2.500 \tag{18b}$$

$$\overline{x^2} = \tfrac{1}{4}\Big((1.00)^2 + (2.00)^2 + (3.00)^2 + (4.00)^2\Big) = 7.500 \tag{18c}$$

$$\overline{y^2} = \tfrac{1}{4}\Big((1.10)^2 + (2.00)^2 + (2.90)^2 + (4.00)^2\Big) = 7.405 \tag{18d}$$

$$\overline{xy} = \tfrac{1}{4}\Big((1.00)(1.10) + (2.00)(2.00) + (3.00)(2.90) + (4.00)(4.00)\Big) = 7.450 \tag{18e}$$

Eqq. (17) then give

$$m = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{7.450 - (2.500)(2.500)}{7.500 - (2.500)^2} = 0.960 \tag{19a}$$

$$b = \frac{\overline{y}\,\overline{x^2} - \overline{x}\,\overline{xy}}{\overline{x^2} - \overline{x}^2} = \frac{(2.500)(7.500) - (2.500)(7.450)}{7.500 - (2.500)^2} = 0.100 \tag{19b}$$

so that the equation of the best-fit line is

$$y = 0.960\,x + 0.100$$

As expected, this is indeed close to $y = x$.

2.2 Errors in Curve Fitting

2.2.1 The Correlation Coefficient

Once a set of data points has been fitted to a line by least squares, there is of course the question how well the data points in fact fit the line. One measure of this is the correlation coefficient, which is defined to be

$$\frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2}\,\sqrt{\overline{y^2} - \overline{y}^2}} \tag{20}$$

Why this is a good measure of how well the data points fit to the line is a little hard to see at our level — in some sense the correlation coefficient is the statistical equivalent of a dot product between unit vectors — , but the following properties of this definition will, we hope, at least make it seem plausible that it provides such a measure:

- The correlation coefficient is dimensionless, so that it does not depend on the units used to measure x and y: if we denote the dimensions of x by $[x]$ and those of y by $[y]$, then the dimensions of the correlation coefficient are
$$\frac{[x][y]}{\sqrt{[x]^2}\,\sqrt{[y]^2}} = 1$$

- When the value of y is independent of the value of x — that is, when there is no correlation between x and y — , then the average value of the product xy should be the same as the product of the average values of x and y, so that

$$\overline{xy} = \overline{x}\,\overline{y}$$

and consequently $\overline{xy} - \overline{x}\,\overline{y} = 0$ and the correlation coefficient vanishes. Although we will not prove the converse, it does turn out that when the correlation coefficient is close to zero, y is largely independent of x and there is little justification for concluding that they are linearly related.

- When the correlation is perfect, so that all of our points lie exactly on the line $y = mx + b$, then

$$\begin{aligned}
\overline{y} &= \overline{mx + b} = m\overline{x} + b \\
\overline{xy} &= \overline{x(mx + b)} = \overline{mx^2 + bx} = m\overline{x^2} + b\overline{x} \\
\overline{y^2} - \overline{y}^2 &= \overline{(mx + b)^2} - (m\overline{x} + b)^2 \\
&= \overline{m^2x^2 + 2mxb + b^2} - (m^2\overline{x}^2 + 2m\overline{x}b + b^2) \\
&= (m^2\overline{x^2} + 2m\overline{x}b + b^2) - (m^2\overline{x}^2 + 2m\overline{x}b + b^2) \\
&= m^2(\overline{x^2} - \overline{x}^2)
\end{aligned}$$

and the correlation coefficient becomes

$$\frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2}\,\sqrt{\overline{y^2} - \overline{y}^2}} = \frac{(m\overline{x^2} + b\overline{x}) - \overline{x}(m\overline{x} + b)}{\sqrt{\overline{x^2} - \overline{x}^2}\sqrt{m^2(\overline{x^2} - \overline{x}^2)}} = \frac{m(\overline{x^2} - \overline{x}^2)}{\sqrt{m^2(\overline{x^2} - \overline{x}^2)^2}}$$

If you look back at our derivation of eq. (7), you will see that quantities like $\overline{x^2} - \overline{x}^2$ are always positive. But since the slope m can be either positive or negative, the correlation coefficient reduces to [32]

$$\frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2}\,\sqrt{\overline{y^2} - \overline{y}^2}} = \frac{m}{\sqrt{m^2}} = \pm 1$$

Although we will not prove the converse, it does turn out that when the correlation coefficient is close to ± 1, y is strongly linearly related to x.

For the sample data on p.30, we had already worked out in eqq. (18) the various required averages; using these in eq. (20) yields as the correlation coefficient of our least-squares fit

$$\frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2}\,\sqrt{\overline{y^2} - \overline{y}^2}} = \frac{7.450 - (2.500)(2.500)}{\sqrt{7.500 - (2.500)^2}\,\sqrt{7.405 - (2.500)^2}} = 0.999 \qquad (21)$$

This is indeed very close to 1, indicating that, as expected, the data points fit very well to a straight line.

[32] If $m = 0$, $y = mx + b$ reduces to $y = b$, so that y is independent of x — which doesn't really fall among the cases of perfect correlation we are considering.

2.2.2 χ^2 & Standard Error

Another question that arises after performing a least-squares fit is how accurate the results for the slope and intercept are. As a first step toward determining their accuracy, we define χ^2_n to be the average squared distance between the line and the n data points:[33]

$$\chi^2_n = \frac{1}{n}\sum_{i=1}^{n}\Big(y_i - (mx_i + b)\Big)^2$$

Since χ^2_n involves exactly the same sum that we minimized in determining the least-squares fit, you might expect that the smaller the χ^2_n, the better the fit. And you would almost be right. The problem is that judging the closeness of the fit by averaging the squared distance over all n data points can sometimes be misleading. If, for example, we had only two data points, then the best-fit line would pass directly through both points and the χ^2_n would vanish. But to conclude from $\chi^2_n = 0$ that the data fit perfectly to a straight line would be silly: since it takes two points to determine a straight line, the data can't even begin to show a deviation from a straight line until we have at least three data points. More generally, since it takes two points to determine a straight line, out of n data points there are only, in effect, $n-2$ points capable of deviating from the line. These $n-2$ points are thus known as the number of degrees of freedom remaining in the data: two of the original n degrees of freedom are eaten up in determining the two parameters of the fit (the slope and intercept), so that only $n-2$ remain. A better measure of the closeness of fit would therefore be what is sometimes called the reduced χ^2 and denoted by χ^2_{n-2}:

$$\chi^2_{n-2} = \frac{1}{n-2}\sum_{i=1}^{n}\Big(y_i - (mx_i + b)\Big)^2$$

We can in fact regard χ^2_{n-2} as the error in the fit, in the sense of a sort of standard deviation.[34]

Our next task is to relate this error in the fit as a whole to the error in the individual parameters of the fit — in the case of a straight line, to the errors in the slope and the intercept. To do this will require us to delve into some linear algebra, and if we're going to do that, we might as well do the derivation for a more general case than a linear fit — it won't be any more or less painful or hard to follow. Just fasten your seat belts and, if you want, close your eyes.

Suppose we expect the n data points (x_i, y_i), $i = 1, \ldots, n$, to fit a functional relationship $y = f(x)$, where the function f involves p parameters c_j, $j = 1, \ldots, p$. The change δy_i in value of y corresponding to x_i when we vary a

[33] χ is the lowercase Greek letter *chi*, and χ^2 is pronounced like *sky squared* without the initial *s*. Just so you don't end up embarrassing yourself at a cocktail party by talking about that "squiggly little cross thingie that's squared."

[34] For fits to curves involving p parameters, $n-p$ degrees of freedom remain among the n data points, and χ^2 would of course be averaged over $n-p$ rather than over $n-2$.

particular parameter c_j by δc_j will be

$$\delta y_i = \frac{\partial f(x_i)}{\partial c_j}\,\delta c_j$$

The total change δy_i when we vary all of the p parameters c_j, each by its own amount δc_j, will be the sum of the changes due to each parameter individually:

$$\delta y_i = \sum_{j=1}^{p} \frac{\partial f(x_i)}{\partial c_j}\,\delta c_j$$

This total change can be conveniently expressed in matrix form as

$$\delta \mathbf{y} = \mathbf{M}\,\delta \mathbf{c} \tag{22}$$

where the elements of the column vectors $\mathbf{y}$ and $\mathbf{c}$ are the y_i and c_j, respectively, and the matrix $\mathbf{M}$ is

$$\mathbf{M} = \begin{pmatrix} \dfrac{\partial f(x_1)}{\partial c_1} & \dfrac{\partial f(x_1)}{\partial c_2} & \cdots \\ \dfrac{\partial f(x_2)}{\partial c_1} & \dfrac{\partial f(x_2)}{\partial c_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

that is, $\mathbf{M}$ is an $n \times p$ matrix with elements

$$M_{ij} = \frac{\partial f(x_i)}{\partial c_j}$$

With this matrix notation, the sum of the squared changes δy_i is

$$\begin{aligned} \sum_{i=1}^{n} (\delta y_i)^2 &= \sum_{i=1}^{n} (\delta y_i)(\delta y_i) \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} M_{ij}\,\delta c_j \right) \left(\sum_{j'=1}^{m} M_{ij'}\,\delta c_{j'} \right) \\ &= \sum_{j,j'=1}^{p} \left(\sum_{i=1}^{n} M_{ij} M_{ij'} \right) \delta c_j\,\delta c_{j'} \\ &= \sum_{j,j'=1}^{p} \left(\sum_{i=1}^{n} M^T_{ji} M_{ij'} \right) \delta c_j\,\delta c_{j'} \\ &= \sum_{j,j'=1}^{p} (M^T M)_{jj'}\,\delta c_j\,\delta c_{j'} \end{aligned}$$

where $\mathbf{M}^T$ is the transpose of the matrix $\mathbf{M}$. Multiplying both sides by $(\mathbf{M}^T\mathbf{M})^{-1}$, we obtain

$$\delta c_j\,\delta c_{j'} = (\mathbf{M}^T\mathbf{M})^{-1}_{jj'} \sum_{i=1}^{n} (\delta y_i)^2$$

which, in the special case $j' = j$, reduces to

$$(\delta c_j)^2 = (\mathbf{M}^T\mathbf{M})^{-1}_{jj} \sum_{i=1}^{n} (\delta y_i)^2$$

If we now think in terms of errors rather than differential variations, we can identify

$$\sum_{i=1}^{n} (\delta y_i)^2$$

with the reduced chi squared χ^2_{n-p} and identify the δc_j with the error in the parameter c_j — which, since it is sort of like a standard deviation, we will denote by σ_j. We thus arrive at

$$\sigma_j^2 = \chi^2_{n-p} \, (\mathbf{M}^T\mathbf{M})^{-1}_{jj} \tag{23}$$

Eq. (23) is the general relation for the standard error σ_j in the parameter c_j.

Our final task is to work out this standard error for the special case of a linear fit. For such a fit, the two parameters are $c_1 = m$ and $c_2 = b$, with $f(x) = mx + b$. Since

$$\frac{\partial f(x_i)}{\partial m} = \frac{\partial}{\partial m}(mx_i + b) = x_i \qquad \frac{\partial f(x_i)}{\partial b} = \frac{\partial}{\partial b}(mx_i + b) = 1$$

the matrix $\mathbf{M}$ is

$$\mathbf{M} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \end{pmatrix}$$

Doing out the matrix multiplication, we have

$$(\mathbf{M}^T\mathbf{M})^{-1} = \left(\begin{pmatrix} x_1 & x_2 & x_3 & \cdots \\ 1 & 1 & 1 & \cdots \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \end{pmatrix} \right)^{-1} = \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} 1 \end{pmatrix}^{-1}$$

which, if we note that

$$\frac{1}{n}\, n = \frac{1}{n}\left(\frac{1}{n}\right)^{-1}$$

may be gerrymandered into

$$(\mathbf{M}^T\mathbf{M})^{-1} = \frac{1}{n} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_i^2 & \frac{1}{n}\sum_{i=1}^{n} x_i \\ \frac{1}{n}\sum_{i=1}^{n} x_i & \frac{1}{n}\sum_{i=1}^{n} 1 \end{pmatrix}^{-1} = \frac{1}{n} \begin{pmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{pmatrix}^{-1}$$

Evaluating the inverse matrix by the usual methods yields

$$\begin{pmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{pmatrix}^{-1} = \frac{1}{\overline{x^2} - \overline{x}^2} \begin{pmatrix} 1 & -\overline{x} \\ -\overline{x} & \overline{x^2} \end{pmatrix}$$

so that we arrive at

$$(\mathbf{M}^T\mathbf{M})^{-1} = \frac{1}{n\Big(\overline{x^2} - \overline{x}^2\Big)} \begin{pmatrix} 1 & -\overline{x} \\ -\overline{x} & \overline{x^2} \end{pmatrix}$$

By eq. (23), the standard error σ_m in the slope, which corresponds to the 11 element, is therefore given by

$$\sigma_m^2 = \frac{1}{n\Big(\overline{x^2} - \overline{x}^2\Big)}\, \chi^2_{n-2} \tag{24a}$$

and the standard error σ_b in the intercept, which corresponds to the 22 element, by

$$\sigma_b^2 = \frac{\overline{x^2}}{n\Big(\overline{x^2} - \overline{x}^2\Big)}\, \chi^2_{n-2} \tag{24b}$$

As an example of the calculation of standard errors for a least-squares fit, recall now the data points of p.30:

$$(1.00, 1.10) \qquad (2.00, 2.00) \qquad (3.00, 2.90) \qquad (4.00, 4.00)$$

To apply eqq. (24), we need first to calculate the reduced χ^2. Using our previously determined slope and intercept from eqq. (19), we have

$$\begin{aligned} \chi^2_{n-2} &= \frac{1}{n-2} \sum_{i=1}^{n} (\delta y_i)^2 \\ &= \frac{1}{n-2} \sum_{i=1}^{n} \Big(y_i - (mx_i + b)\Big)^2 \\ &= \frac{1}{4-2} \sum_{i=1}^{4} \Big(y_i - (0.9600\, x_i + 0.1000)\Big)^2 \end{aligned}$$

which, when we plug the four data points (x_i, y_i) into the sum, yields [35]

$$\chi^2_{n-2} = \tfrac{1}{2}\Big((0.040)^2 + (0.020)^2 + (-0.080)^2 + (-0.060)^2\Big) = 0.00600$$

By eqq. (24) and our previous results $\overline{x^2} = 7.500$ and $\overline{x} = 2.500$ from eqq. (18), the standard errors in m and b are then

$$\sigma_m^2 = \frac{1}{n\Big(\overline{x^2} - \overline{x}^2\Big)}\, \chi^2_{n-2} = \frac{1}{4(7.500 - (2.500)^2)}\, 0.00600 = 0.00120$$

[35] As usual, we are taking some liberties with significant figures in this error calculation.

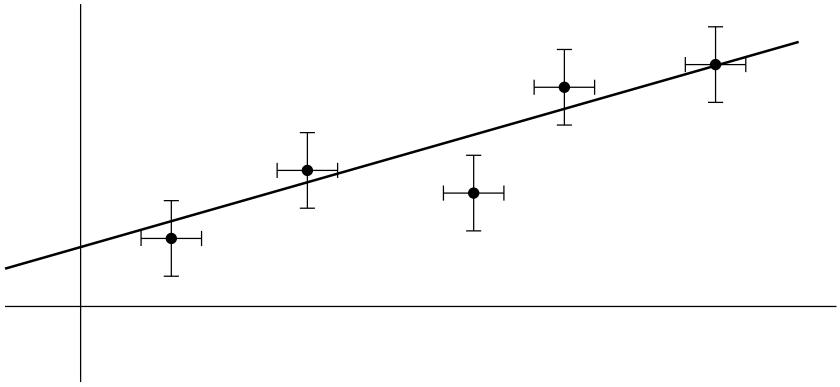

Figure 4: A Plot with Error Bars. Viewer Discretion is Advised.

$$\sigma_b^2 = \frac{\overline{x^2}}{n\left(\overline{x^2} - \overline{x}^2\right)} \chi^2_{n-2} = \frac{7.500}{4(7.500 - (2.500)^2)} 0.00600 = 0.00900$$

whence

$$\sigma_m = 0.035 \qquad\qquad \sigma_b = 0.095$$

Putting this together with our results from the previous sections, the best fit is therefore the line $y = mx + b$ with

$$m = 0.960 \pm 0.035 \qquad\qquad b = 0.100 \pm 0.095$$

and a correlation coefficient of 0.999.

2.3 Error Bars

An error bar is a line drawn through a data point to indicate visually the amount of error in its ordinate or abscissa. Suppose, for example, you are plotting the magnitude of the spring force F versus the distention x of a spring and that your values of F and x in the data point (3.2 cm, 1.5 N) are accurate to ± 0.1 cm and ± 0.2 N, respectively. To show the error in x, you would draw a horizontal bar through this data point extending from $x = 3.2 - 0.1 = 3.1$ cm to $x = 3.2 + 0.1 = 3.3$ cm. Likewise, to show the error in F, you would draw a vertical bar through this data point extending from $F = 1.5 - 0.2 = 1.3$ N to $F = 1.5 + 0.2 = 1.7$ N. Fig. (4) shows what a plot with error bars looks like. The extent of the error bars need not, as it is in fig. (4), be the same for all points, nor is it necessary to have the little feet on your bars.[36] And, like errors themselves, error bars can represent confidence levels.

While the least-squares fit, the standard errors, and the correlation coefficient are all you need to assess quantitatively how well your data fits to a straight line and how much error there is in your slope and intercept, error

[36] Though you have to admit they do look pretty snazzy, if we say so ourselves.

bars are a useful aid in visualizing how well your line fits your individual data points: you can see whether your line passes through the rectangle formed by the horizontal and vertical error bars of each point.

3 The Lab Report

3.1 The Lab Notebook

In a word: you don't need one.[37] You can use a lab notebook if you want, but most people find it easier to write the lab up using a word processor and importing tables and graphs into the main document from various other applications. It's also perfectly acceptable to write the lab up by hand — as long as it's neat and legible.

3.2 Feel Free to Use the First Person ...

In the old days, the convention was to write scientific papers entirely in the passive voice. This was one of many silly things people did in the old days. While some still use the passive voice, the vast majority of physicists use the first person nowadays. You should therefore not hesitate to use *I* or *we* where it is natural to do so.

3.3 ... But Do Speak English

If we were going to adhere to tradition, we would also be making you write your lab reports in Latin, but it seems that most students find writing in English more than sufficiently challenging. Even the native speakers of English. In fact, foreign students often write better English than American students.

But wherever you're from, you want to be very clear, precise, and concise. Nothing irrelevant should be included, nothing relevant omitted. You may assume familiarity with the physics, but a reader unfamiliar with the equipment and procedure should be able to follow what you did in the lab and with the data clearly enough to reproduce your work on his or her own. To this end,

- Be sure that your report is consistent with a healthy degree of self-respect. There is no excuse for measuring with a *rular* rather than a *ruler*. Or with a *veneer calibrator* or a *Veneal calibur* instead of a *Vernier caliper*.

- Say what you mean. It will likely require some struggling to find the right words, but that this is a science course, not an English course, is all the more reason to be clear and precise. No lab report should state that "a few errors resulted in the lab" or speak of "opportunities for error."

[37] The astute reader will have observed that that's four words, but who's counting?

- Avoid verbosity and vacuous observations. After you have written a sentence or a paragraph, reread it to make sure that you have said something substantial, and said it not only clearly and precisely, but in the fewest possible words.

Consider, for example, these quotes from lab reports we have received:

> "One of the most easily made errors that could have been made throughout the lab would be in measurements."
>
> "Numerous measurements were made that would have significantly effected [*sic*] our calculations."
>
> "Because most of this lab was done by lab equipment, there was not much room for error."
>
> "Inaccuracies in measurement can account for much of our experimental error."
>
> "The consideration of the effect of timing-related error prevents the resulting value of g to be unduly influenced."
>
> "A few errors became apparent in the percent-error calculation."
>
> "...causing an inflammation of the recorded times ...."
>
> "Another possible error that could have occurred as in any lab is the possibility of misreading measurements."

You don't improve your writing by writing. You improve your writing by determined, painstaking rewriting. The alternative is to end up being quoted in this section.

3.4 Anatomy of a Lab Report

3.4.1 The Title, &c.

Every lab report should have a title followed by a clear indication of the date of the report, the author's name,[38] and, equally importantly, the names of your lab partners.

3.4.2 The Abstract

Even within specialized areas of physics, such as high-energy theory or quantum cosmology, there are usually scores of papers published every day — thousands every year. Researchers trying to keep up with their fields or to learn more about a specific topic do not have time to sift through all of these.

[38] Dude: that means your name.

Every paper therefore begins with an abstract that briefly summarizes what the paper is about: by glancing at the abstract, a researcher can quickly ascertain whether the body of the paper will be of enough relevance and interest to be worth reading.

To serve its purpose well, an abstract should indicate

- The nature of the experiment or, in the case of a paper in theoretical physics, of the calculation.
- Anything especially significant about equipment or methods.
- The results obtained, including, for experimental work, the error and confidence level.
- The significance of the results, including consistency or inconsistency with any relevant previously published results.

Though an abstract must be clear and complete, succinctness is a paramount virtue; there should be no unnecessary details. Often a few sentences suffice. Lack of brevity will defeat the purpose of an abstract and is by far the most common beginner's error.

For example, a very lean abstract for a lab in which the acceleration g due to gravity was determined by timing the swings of a pendulum might read

Abstract

> By timing the swings of a simple pendulum of measured length, we determined the acceleration due to gravity at Gassy Swamp, New Jersey to be $g = 9.823 \pm 0.666\ \mathrm{m/s^2}$, consistent with the generally reported value 9.802.

As you can see, this abstract includes everything in the list itemized above. Note also that there is no mention of the instruments or methods used to measure the length of the pendulum or time the swings, because there is nothing especially significant about these. Nor is there any mention of the use of the pendulum relation $g = 4\pi^2\ell/T^2$: this is general knowledge and therefore also not significant enough to merit mention in the abstract.

The abstract should of course precede the body of the paper: burying an abstract elsewhere in the paper, so that the reader has to flip pages to find it, again defeats its purpose.

Although the labs you will be doing in your coursework have a pedagogical purpose, that pedagogical purpose has no place in your lab report: both in the abstract and elsewhere in the report, you should be discussing what you did and what you got, not the teleology of why you did it. Researchers not infrequently do scientific research for all sorts of reasons that have nothing to do with the research itself — because they need to publish to gain tenure or future grants, because they prefer sitting in front of a computer to talking to people, because they enjoy tinkering with electronics, because they get their

jollies torturing rats, or any of infinitely many other reasons — none of which is relevant to the scientific significance or validity of their work. The pedagogical purpose of an experiment in which you determined the acceleration g due to gravity might have been to learn how to perform an error analysis, but that is irrelevant to what you did and what you got, which was to make the requisite measurements and derive from them a result for g.

For some real-world examples of abstracts, see Appendix A.

3.4.3 The Procedure

You should describe the equipment you used and your methods — what you did and how you did it. And as you do this, remember that the purpose is to enable a reader unfamiliar with your equipment and methods to understand your work well enough to reproduce it. All details that serve this purpose should be included, all details that do not should be omitted. As elsewhere in the report, clarity and precision are paramount virtues, but so is concision: you need not write voluminously.

Even when you are provided with a handout that specifies some or all of the equipment and procedure, your report should stand alone: you should not assume that the reader has seen or has access to the handout.

If, for example, the time it takes a cart to pass by an infrared sensor is recorded by a computer, you should certainly indicate this in your report, and probably even note more specifically that the time recorded was the interval during which the infrared beam was broken by the cart, but you should not report irrelevant details of the software, such as hitting the 'start' button before each trial. Similarly it goes without saying that you turned the computer on, plugged the apparatus in, and recorded the times. *Duh!*

Note also that calculations are *not* a part of your experimental procedure, so you should not be including narratives like "we solved the quadratic projectile relation for the time of flight" — you simply show the relations and the calculation.

3.4.4 Data & Tables

The results of all measurements you made should be clearly recorded and labeled in your report: never omit raw data.[39] If, for example, you timed for each of a gazillion trials how long it took an object to travel a fixed distance, all gazillion time measurements should be included in your report — even if you use only the average and standard error of these times in your calculations.

And of course don't forget to append units to every quantity that has a physical dimension. Neglecting to do so not only renders your work meaningless, but could result in a savage beating. Likewise all numerical quantities

[39]The one exception to this would be a set of data taken incorrectly — say, for example, a series of times when you forgot to reset the timing mechanism before each trial. In cases of such clear mistakes, the bad data may simply be discarded.

Distance (cm)	Time (s)	Avg. Speed (cm/s)
32.1	4.56	7.039
21.3	5.64	3.777
13.2	6.45	2.047
12.3	6.54	1.881
31.2	5.46	5.714
23.1	4.65	4.968

Table 2: Distance, Time, & Average Speed

should be expressed with the correct number of significant figures.

When you are making a repeated measurement, there may be trials when the value of the measurement is so anomalous that something clearly went wrong. The determination of such cases is to some degree subjective, but the data points in question should be very clearly pathological. You might, for example, have taken several dozen measurements of the timing of a particular pendulum's swing and found that these ranged from 1.234 to 1.345 sec, except for one value of 8.901 sec. Even if you can't figure out the cause, clearly something went wrong during the trial that gave this exceptional measurement. In such a case, you should still include the anomalous value with your other raw data, but you should note that you are excluding it as an anomaly.

Often your data or the results derived from it can be most clearly presented in a table. If, for example, you measured the distance an object traveled and the corresponding time interval for each of several trials and then used this data to calculate the object's average speed, your table might look like table (2). In fact, in certain respects it *should* look like table (2):

- Every table should have a caption.
- Units should appear only once, at the top of each column. Putting units on each numerical entry would clutter the table, making it hard to read.
- Usually there is a horizontal rule separating the headings from the data, and there may or may not be vertical rules separating columns, but there should never be horizontal rules between data lines. People who do this would wear plaids with checks and argyle socks.

If you break a table across pages because it has more entries than will fit on one page, you should repeat the headings on each page.[40] Alternatively, if the total width of the columns is narrow enough, you might repeat the columns side-by-side.

[40] In sensible typesetting applications, this is very easy to do. If you are using software produced by evil monopolists and have difficulty with this, that's your problem.

3.4.5 Calculations

The reader should also be able to follow and reproduce your calculations. While you need not show the calculation of statistical quantities such as means or standard errors — such calculations can be assumed general knowledge — , you should clearly explain all calculations specific to your experiment:

- Explicitly indicate the general relations on which you are basing your calculations.
- Indicate what each variable in each relation represents.
- Explicitly indicate your numerical substitutions.
- Show the steps of any nontrivial calculations.

If, for example, you are calculating the acceleration g due to gravity from the pendulum relation

$$T = 2\pi\sqrt{\frac{g}{\ell}}$$

your report should include some variation of

> Solving the pendulum relation
>
> $$T = 2\pi\sqrt{\frac{g}{\ell}}$$
>
> for the acceleration g due to gravity, we obtain
>
> $$g = \frac{4\pi^2\ell}{T^2}$$
>
> where ℓ and T are the length and period of the pendulum, respectively. With our measured value of 0.995 m for the length and our mean value of 2.00468 sec for the period, this yields
>
> $$g = \frac{4\pi^2\ell}{T^2} = \frac{4\pi^2 \times 0.995}{2.00468^2} = 9.77\ \mathrm{m/s^2}$$
>
> blah, blah, blah, yada, yada, yada ....

Again, be sure to pay attention to significant figures when you do your calculations.

3.4.6 Graphs

How lines are fitted to sets of data points was discussed in §2; in this section we are concerned only with matters of formatting.

Bigger is better. A graph is a visual representation of data. Since visual representations that can't be clearly made out are of very little use, you should make your graphs *big* — big enough that your data points, the curve through them, and all the other flora and fauna of the plot are clearly discernible. Plots that fill the whole page, or at least some substantial fraction of the page, are best.[41]

Grids are good. In the old days, you would be doing plots by hand on graph paper. Although these days everyone simply hits the 'print' button instead, it is often still highly desirable that your plot have a grid, because it makes it much easier to discern the quantitative values of your data points.[42] You should choose the size of your grid boxes with this purpose in mind: if too large, the grid won't serve its purpose, but if too small, it will clutter the graph and make it hard to read.

Label the axes. Every graph should have a caption, and you should clearly indicate on each axis the quantity being graphed, the numerical scale, and the units. The numerical values of the demarcations (the grid boxes) on the axes should be labeled finely enough to make the values of your data points easily discernible, but not so finely that they clutter the graph and make it hard to read.

Choose the scale wisely. Set the scale of the axes so that the data points are widely spread across the graph, bearing in mind that axes need not start at zero. A big graph does not have much virtue if all of your data points are smashed together in one corner.[43]

Don't embarrass yourself. Even if the curve that fits your data points is unknown, you should never succumb to the temptation to play "connect-the-dots" by drawing straight line segments between adjacent data points.[44] Unless doing so would produce a humorous caricature of Donald Duck and this is somehow relevant to your analysis.

[41] If you are using a graphing application that habitually produces pathetically small graphs, then you need to learn how to make it produce larger graphs — or find a better graphing application. Applications produced by evil monopolists seem particularly ill suited in this respect.

[42] If you are using a graphing application that habitually produces graphs without a grid, then you need to learn how to make it produce graphs with grids or find a better graphing application. Applications produced by evil monopolists seem particularly ill suited in this respect.

[43] If you are using a graphing application that insists on starting axes at zero, then you need to learn how to make it do otherwise or find a better graphing application. Applications produced by evil monopolists seem particularly ill suited in this respect.

[44] If you are using a graphing application that likes to play connect-the-dots, then you need to learn how to stop it from doing that or find a better graphing application. Applications produced by evil monopolists seem particularly inclined to play connect-the-dots.

Fig. (5) on the next page shows a respectable plot, replete with variably sized error bars and best-fit line, albeit of an admittedly rather pedestrian and heinously sparse set of data points.

3.4.7 Error Analysis

Performing a rigorous error analysis can be a royal pain, but without such an analysis you cannot assess the validity and import of your quantitative results. A thorough analysis of errors, as discussed in §1, is therefore an essential part of a lab report. Suck it up.

3.4.8 Discussion

As discussed on p.40 in §3.4.2, the pedagogical purpose of the lab is irrelevant to the science of your lab work and therefore has no place in your report. This is, however, a high-school physics class, not the nuclear weapons lab at Lawrence Livermore:[45] the reality behind the facade[46] is that our labs are in fact pedagogical exercises, and you must therefore demonstrate an understanding of the physics of the experiment, of error analysis, and of the import of your results. You should devote to this discussion time and energy commensurate with its importance.

There is no formulaic way to determine the points you should discuss; these vary not only from lab to lab but from group to group: different results raise different issues. In general, however, you should fully discuss

- The methods by which your measurements were made.
- All sources of error.
- The reasoning behind your estimates of quantifiable errors.
- The nature of any unquantifiable errors and, if possible, of the effect they would have had on your measurements and results.
- Any assumptions or approximations you made.
- The import of your results, including consistency or inconsistency with any other reported results.
- Any conclusions you draw.
- Any observations warranted by the data, equipment, methods, results, or other aspects of the lab, such as a trend in your data, a limitation of the equipment, an anomalous data point, etc., etc.

[45] As those of you who keep up with the news are aware, the principal difference between nuclear-weapons research facilities and high-school physics labs is that the high-school labs have tighter security.

[46] Or facade behind the reality. Take your pick.

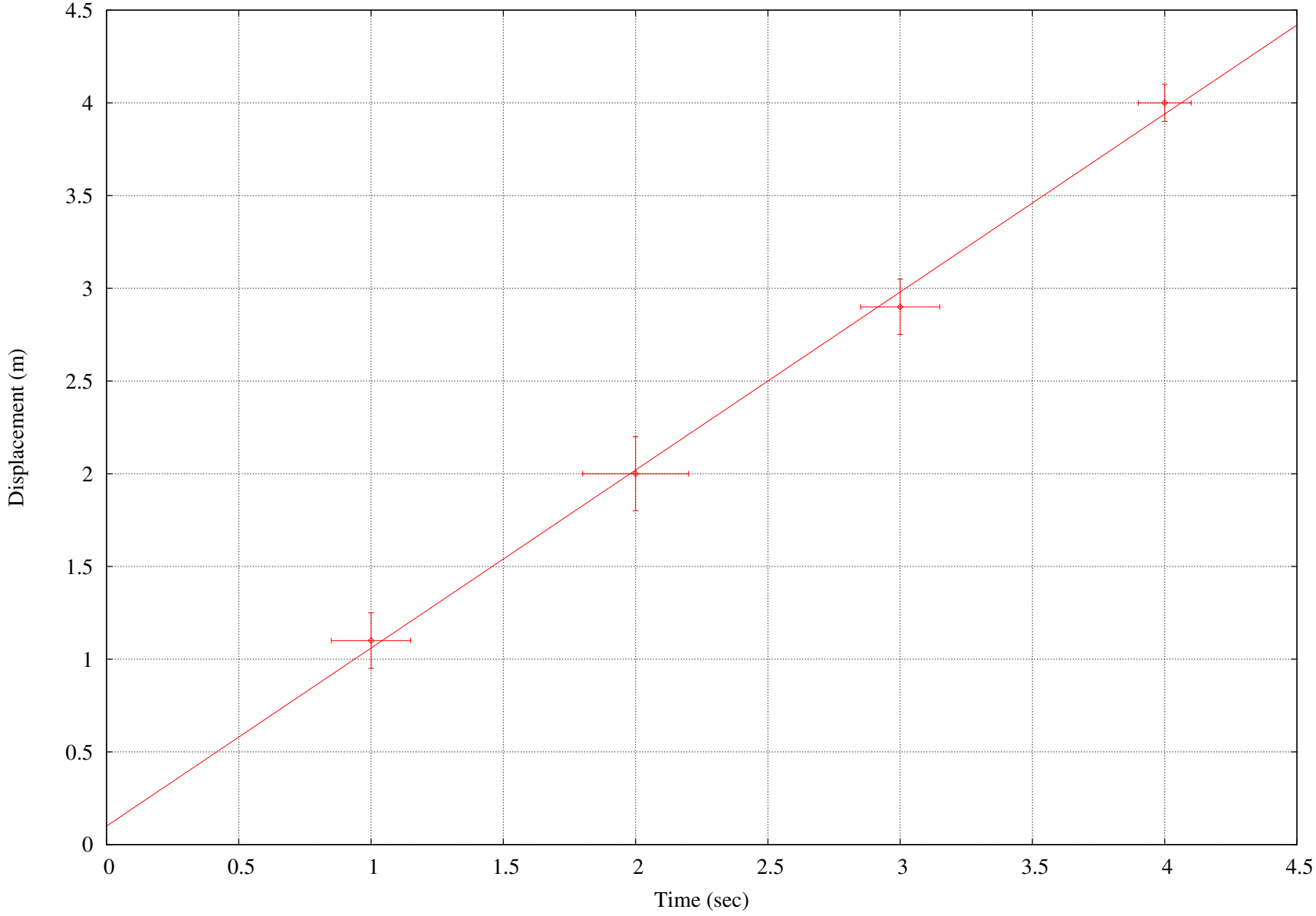

Figure 5: Displacement versus Time

If, for example, your results for the velocity of an object all came out higher than expected, you should note this and thoroughly discuss what you can or cannot surmise about this effect and its cause: Is the deviation even significant, and if so, could it be purely random, or would it likely be systematic? Could a specific source of error that was not taken into account be responsible? Or a quirk or limitation of the method or equipment?

Whatever point you are discussing, *be explicit*, even if the point to be made seems self-evident. Do not assume anything goes without saying. Later on, when you are doing real research in academia or industry, you can try to make yourself look smart by being obscure, but for your present pedagogical exercises, what isn't on the page doesn't do much to demonstrate your understanding.

Whatever you discuss should, however, be relevant and significant. While it is often not only acceptable but desirable to discuss effects such as friction even when you cannot quantify them and may not even be sure that they were of significant magnitude, you should not be resorting to fantastic speculation in the mistaken belief that the reader will be impressed with your comprehensiveness. A longer report is not necessarily better; when there is no more to be said, there is no more to be said. Enough said.

3.4.9 Conclusion

The conclusion always comes at the very end of the report. There wouldn't be much point in calling it a conclusion otherwise, now would there?

In some ways, a conclusion is superfluous: any conclusions you draw should already have been thoroughly discussed and rigorously justified — quantitatively whenever possible — in the body of your report. The formal conclusion is merely a concise reiteration of those conclusions, and as such it is often very similar to the abstract not only in form but in content. If the abstract for a report read

Abstract

> By timing the swings of a simple pendulum of measured length, we determined the acceleration due to gravity at Gassy Swamp, New Jersey to be $g = 9.823 \pm 0.666\ \mathrm{m/s^2}$, consistent with the generally reported value 9.802.

then the conclusion might read

> By timing the swings of a simple pendulum of measured length, we determined the acceleration due to gravity at Gassy Swamp, New Jersey to be $g = 9.823 \pm 0.666\ \mathrm{m/s^2}$, consistent with the generally reported value 9.802.

Very likely, however, the conclusion section would also include conclusions reached about the significance or insignificance of friction or other noteworthy

aspects of the data or results beyond the scope of the abstract — so you shouldn't be lazy and simply repeat your abstract.

You also shouldn't succumb to the temptation to make your conclusion sound like an epic on human struggle and its ultimate triumph.[47] In an experiment that determined the value of the acceleration g due to gravity, you should not be noting that you succeeded in measuring g or that you accomplished the purpose of the lab by determining g. You determined g. Period. The human drama is irrelevant.

3.4.10 Putting the Pieces Together

Now that we've discussed the various parts of the lab report separately, there is the question of how to combine them into a coherent whole. The conventional prescription is very simple:

- The abstract always precedes the body of the report.
- The conclusion always comes at the very end of the report.
- The other parts of the report may be presented in any manner that is natural, logical, clear, and consistent with §3.4.1 through §3.4.9.

As long as you adhere to these three requirements, the organization of the report, and in particular the degree to which its various parts are separate or intertwined, is up to you.

3.5 Outline of a Sample Lab Report

In this section we *outline* the elements of the report one might write for an experiment that determined the acceleration g due to gravity by timing the swings of a simple pendulum.[48]

The report would of course begin with a title with all the elements discussed in §3.4.1. Immediately following this would be the abstract, which might read

Abstract

> By timing the swings of a simple pendulum of measured length, we determined the acceleration due to gravity at Gassy Swamp, New Jersey to be $g = 9.823 \pm 0.666\ \mathrm{m/s^2}$, consistent with the generally reported value 9.802.

The report would then describe the relevant aspects equipment and procedure:[49]

[47] Or failure, depending on your results.

[48] Those of you who were expecting a full writeup may want to check the definition of the words *entitlement* and *lazy* in the dictionary. Though we'll admit that it is unclear to which of us the latter more applies.

[49] We are being inconsistent here about the use of the use of the first person and the passive voice, so that you have examples of both styles.

A pendulum was constructed by securing one end of a thin string to a pivot and the other end to the center of a cylindrical mass. To minimize friction, the pivot end of the string was secured by a knot that remained immobile as the pendulum swung. The attached mass was far greater than the mass of the string and of an extent far less than the length of the string.

From the knot at the pivot to the center of the mass, the length of the string, when measured with a standard meter stick, was determined to be 0.995 m. The size of the suspended mass, about 2 cm, was judged small enough for the mass to be treated as a point mass.

The pendulum was set in motion and 100 swings were timed for each of six trials with a stopwatch accurate to $\frac{1}{100}$ sec.

The report would then go on to present a table of the raw time data for the six trials; the estimate of the error in the length measurement, including a full, quantitative discussion of the sources of this error and its estimation; and the results for the mean and standard error in the time measurement, as discussed in §1.6 and §1.8. The general relation for a simple pendulum would be solved for g, and the calculation of the result for g from the measurements of the pendulum's length and its period would be shown.

The calculation of the mean and standard error of the time measurement can be assumed general knowledge and need not be shown, but the reasons for estimating the error in the time measurement statistically — variability that could not be directly estimated, both in starting and stopping the stopwatch and in judging when the pendulum had reached the far point of its swing — would be explained, and the use of the relevant error-propagation relation to determine the error in g would be presented in detail.[50] The final result for g would be quoted, along with its absolute error, units, and confidence level.

At the very least, the report would also discuss the effect of friction. It would be noted that there was certainly some friction in the apparatus, including air resistance to the motion of the string and bob, and that the effect of this air resistance would be to retard the motion and thus elongate the time measurement. Explicitly reasoning from the relation for g, the report would show that friction would therefore have tended to depress the result for g. It would, however, be noted that the magnitude of the effect of friction, and even whether its magnitude was significant, was not known. But any qualified conclusions that could be made about the effect of friction would be thoroughly reasoned out, as discussed in §1.10.

A really snazzy report might also reason out the effect of the bob being, not a point mass, but of finite extent.

At the end of the report would be the conclusion, which, in addition to

[50]The derivation of the error-propagation relation itself need not be explained when it requires the application of eq. (1) and thus multi-variable calculus. In these cases, the error-propagation relation will be presented in class.

largely reiterating the abstract, would also succinctly summarize any conclusions reached about frictional or other effects discussed in the body of the report.

A Examples of Abstracts

You probably won't understand any of the physics to which these abstracts refer, but they should give you some idea of the kinds of things that an abstract should include.

Note that these abstracts, while very precise and complete, have no fat on them. Lean and mean is what you want.

Measurement of the Helicity of W Bosons in Top-Quark Decays
A. Abulencia, et al. (CDF Collaboration)

> We measure the branching fraction of the top quark to longitudinally and right-handed polarized W bosons, F_0 and F_+, using approximately $200\,\mathrm{pb}^{-1}$ of $\bar{p}p$ collisions collected by the CDF II detector. We analyze two quantities sensitive to the W helicity: the invariant mass of the charged lepton and the bottom-quark jet in the decay $t \to Wb \to \ell\nu b$ (where $\ell = e$ or μ), and the transverse momentum of the charged lepton. We find $F_0 = 0.74^{+0.22}_{-0.34}$, and $F_+ < 0.27$ at the 95% confidence level. These measurements are in agreement with the standard model predictions.

Measurement of the WW Cross Section in 7 TeV pp Collisions with the ATLAS Detector
Haijun Yang

> We report a measurement of the WW production cross section in pp collisions at 7 TeV. The WW leptonic decay channels are analyzed using data corresponding to 1.02 /fb of integrated luminosity collected by the ATLAS detector during 2011 at the CERN Large Hadron Collider. With 414 observed WW candidate events and an estimated background of 170±28 events, the measured WW production cross section is 48.2±4.0(stat)±6.4(syst)±1.8(lumi) pb, in agreement with the Standard Model NLO prediction of 46±3 pb.

Improved Bounds on Universal Extra Dimensions
Thomas Flacke

We report on recent constraints on models with a flat "universal" extra dimension in which all Standard Model fields propagate in the bulk. A significantly improved constraint on the compactification scale is obtained from the extended set of electroweak precision observables accurately measured at LEP1 and LEP2. We find a lower bound of $M_c \equiv R^{-1} > 700$ (800) GeV at the 99% (95%) confidence level. Comparison of this constraint with the relic density of Kaluza-Klein dark matter for the Minimal UED model points towards the necessity of including non-minimal boundary terms which motivates studying alternative Kaluza-Klein dark matter candidates. Results for the one-loop induced magnetic dipole moment for Kaluza-Klein neutrino dark matter are presented.
This talk is based on Phys.Rev.D73:095002,2006 and hep-ph/0601161.

Search for Baryons in the Radiative Penguin Decay $b \to s\gamma$
K. W. Edwards, et al. (CLEO Collaboration)

We have searched for the baryon-containing radiative penguin decays $B^- \to \Lambda\overline{p}\gamma$ and $B^- \to \Sigma^0\overline{p}\gamma$, using a sample of 9.7×10^6 $B\overline{B}$ events collected at the $\Upsilon(4S)$ with the CLEO detector. We find no evidence for either, and set 90% confidence level upper limits of

$$[\mathcal{B}(B^- \to \Lambda\overline{p}\gamma) + 0.3\mathcal{B}(B^- \to \Sigma^0\overline{p}\gamma)]_{E_\gamma > 2.0 \text{ GeV}} < 3.3 \times 10^{-6} ,$$

$$[\mathcal{B}(B^- \to \Sigma^0\overline{p}\gamma) + 0.4\mathcal{B}(B^- \to \Lambda\overline{p}\gamma)]_{E_\gamma > 2.0 \text{ GeV}} < 6.4 \times 10^{-6} .$$

From the latter, we estimate

$$\mathcal{B}(B \to X_s\gamma, X_s \text{ containing baryons})_{E_\gamma > 2.0 \text{ GeV}} < 3.8 \times 10^{-5} .$$

This limit implies upper limits on corrections to CLEO's recent measurement of branching fraction, mean photon energy, and variance in photon energy from $b \to s\gamma$ that are less than half the combined statistical and systematic errors quoted on these quantities.

Rebuttal to a Paper on Wormholes
L. Susskind

> In a recent paper on wormholes (gr-qc/0503097), the author of that paper demonstrated that he didn't know what he was talking about. In this paper I correct the author's naive erroneous misconceptions.

Composite self-similar solutions for relativistic shocks: the transition to cold fluid temperatures
Margaret Pan,[51] Re'em Sari

> The flow resulting from a strong ultrarelativistic shock moving through a stellar envelope with a polytrope-like density profile has been studied analytically and numerically at early times while the fluid temperature is relativistic—that is, just before and just after the shock breaks out of the star. Such a flow should expand and accelerate as its internal energy is converted to bulk kinetic energy; at late enough times, the assumption of relativistic temperatures becomes invalid. Here we present a new self-similar solution for the post-breakout flow when the accelerating fluid has bulk kinetic Lorentz factors much larger than unity but is cooling through p/n of order unity to subrelativistic temperatures. This solution gives a relation between a fluid element's terminal Lorentz factor and that element's Lorentz factor just after it is shocked. Our numerical integrations agree well with the solution. While our solution assumes a planar flow, we show that corrections due to spherical geometry are important only for extremely fast ejecta originating in a region very close to the stellar surface. This region grows if the shock becomes relativistic deeper in the star.

[51] A former student. Other papers by Margaret Pan:
http://arxiv.org/abs/1309.2289,
http://arxiv.org/abs/1206.3972,
http://arxiv.org/abs/1206.3583,
http://arxiv.org/abs/1111.0667,
http://arxiv.org/abs/1009.1874,
http://arxiv.org/abs/1008.4152,
http://arxiv.org/abs/astro-ph/0505176,
http://arxiv.org/abs/astro-ph/0402138,
http://arxiv.org/abs/astro-ph/0302135,
http://arxiv.org/abs/astro-ph/0205511.

Coherent Control of Ultracold Molecule Dynamics in a Magneto-Optical Trap Using Chirped Femtosecond Laser Pulses
Benjamin L. Brown,[52] Alexander J. Dicks, Ian A. Walmsley

We have studied the effects of chirped femtosecond laser pulses on the formation of ultracold molecules in a Rb magneto-optical trap. We have found that application of chirped femtosecond pulses suppressed the formation of $^{85}Rb_2$ and $^{87}Rb_2$ $a^3\Sigma_u^+$ molecules in contrast to comparable non-chirped pulses, cw illumination, and background formation rates. Variation of the amount of chirp indicated that this suppression is coherent in nature, suggesting that coherent control is likely to be useful for manipulating the dynamics of ultracold quantum molecular gases.

Scintillation time dependence and pulse shape discrimination in liquid argon
W. H. Lippincott,[53] K. J. Coakley, D. Gastler, A. Hime, E. Kearns, D. N. McKinsey, J. A. Nikkel, L. C. Stonehill

Using a single-phase liquid argon detector with a signal yield of 4.85 photoelectrons per keV of electronic-equivalent recoil energy (keVee), we measure the scintillation time dependence of both electronic and nuclear recoils in liquid argon down to 5 keVee. We develop two methods of pulse shape discrimination to distinguish between electronic and nuclear recoils. Using one of these methods, we measure a background and statistics-limited level of electronic recoil contamination to be 7.6×10^{-7} between 60 and 128 keV of nuclear recoil energy (keVr) for a nuclear recoil acceptance of 50% with no nuclear recoil-like events above 72 keVr. Finally, we develop a maximum likelihood method of pulse shape discrimination using the measured scintillation time dependence and predict the sensitivity to WIMP-nucleon scattering in three configurations of a liquid argon dark matter detector.

[52] Another former student.

[53] Yet another former student. Other papers by Hugh Lippincott:
http://arxiv.org/abs/1406.5560, http://arxiv.org/abs/1406.4825,
http://arxiv.org/abs/1403.0525, http://arxiv.org/abs/1310.8327,
http://arxiv.org/abs/1307.8080, http://arxiv.org/abs/1306.5675,
http://arxiv.org/abs/1306.2349, http://arxiv.org/abs/1304.6001,
http://arxiv.org/abs/1211.0909, http://arxiv.org/abs/1204.3094,
http://arxiv.org/abs/1111.6192, http://arxiv.org/abs/1111.3260,
http://arxiv.org/abs/1004.0373, http://arxiv.org/abs/0911.5453,
http://arxiv.org/abs/0904.2930, http://arxiv.org/abs/0709.1880,
http://arxiv.org/abs/astro-ph/0702202,
http://arxiv.org/abs/astro-ph/0612108,
http://arxiv.org/abs/nucl-ex/0503006.

Care and feeding of frogs
Margaret Pan, Eugene Chiang

> "Propellers" are features in Saturn's A ring associated with moonlets that open partial gaps. They exhibit non-Keplerian motion (Tiscareno et al. 2010); the longitude residuals of the best-observed propeller, "Blériot," appear consistent with a sinusoid of period ∼4 years. Pan & Chiang (2010) proposed that propeller moonlets librate in "frog resonances" with co-orbiting ring material. By analogy with the restricted three-body problem, they treated the co-orbital material as stationary in the rotating frame and neglected non-co-orbital material. Here we use simple numerical experiments to extend the frog model, including feedback due to the gap's motion, and drag associated with the Lindblad disk torques that cause Type I migration. Because the moonlet creates the gap, we expect the gap centroid to track the moonlet, but only after a time delay $t_{\rm delay}$, the time for a ring particle to travel from conjunction with the moonlet to the end of the gap. We find that frog librations can persist only if $t_{\rm delay}$ exceeds the frog libration period ¶$_{\rm lib}$, and if damping from Lindblad torques balances driving from co-orbital torques. If $t_{\rm delay} \ll P_{\rm lib}$, then the libration amplitude damps to zero. In the case of Blériot, the frog resonance model can reproduce the observed libration period $P_{\rm lib} \simeq 4$ yr. However, our simple feedback prescription suggests that Blériot's $t_{\rm delay} \sim 0.01 P_{\rm lib}$, which is inconsistent with the observed libration amplitude of 260 km. We urge more accurate treatments of feedback to test the assumptions of our toy models.

The Wind of Variable C in M33
Roberta M.. Humphreys, Kris Davidson, Michael Gordon,[54] Kerstin Weis, Birgitta Burggraf, D. J. Bomans, John C. Martin

> We discuss the spectrum of Var C in M33 obtained just before the onset of its current brightening and recent spectra during its present "eruption" or optically thick wind stage. These spectra illustrate the typical LBV transition in apparent spectral type or temperature that characterizes the classical LBV or S Dor-type variability. LBVs are known to have slow, dense winds during their maximum phase. Interestingly, Var C had a slow wind even during its hot, quiescent stage in comparison with the normal hot supergiants with similar temperatures. Its outflow or wind speeds also show very little change between these two states.

[54] You guessed it: a former student. Other papers by Michael Gordon:
`http://arxiv.org/abs/1301.0824`,
`http://arxiv.org/abs/0906.3030`.

Conductance fingerprint of Majorana fermions in the topological Kondo effect
Martin R. Galpin, Andrew K. Mitchell, Jesada Temaismithi,[55] David E. Logan, Benjamin Béri, Nigel R. Cooper

We consider an interacting nanowire/superconductor heterostructure attached to metallic leads. The device is described by an unusual low-energy model involving spin-1 conduction electrons coupled to a nonlocal spin-$\frac{1}{2}$ Kondo impurity built from Majorana fermions. The topological origin of the resulting Kondo effect is manifest in distinctive non-Fermi-liquid (NFL) behavior, and the existence of Majorana fermions in the device is demonstrated unambiguously by distinctive conductance lineshapes. We study the physics of the model in detail, using the numerical renormalization group, perturbative scaling and abelian bosonization. In particular, we calculate the full scaling curves for the differential conductance in AC and DC fields, onto which experimental data should collapse. Scattering t-matrices and thermodynamic quantities are also calculated, recovering asymptotes from conformal field theory. We show that the NFL physics is robust to asymmetric Majorana-lead couplings, and here we uncover a duality between strong and weak coupling. The NFL behavior is understood physically in terms of competing Kondo effects. The resulting frustration is relieved by inter-Majorana coupling which generates a second crossover to a regular Fermi liquid.

[55] Also a former student.

B Proof of the Central Limit Theorem

The following skeletal outline of a proof of the central limit theorem, presented here only for the sake of completeness, is likely to read like an income-tax form unless you already have some familiarity with probability density functions, normal distributions, and other concepts in statistics.

Probabilities conventionally range from zero to unity, with zero corresponding to impossibility and unity to certainty. If there are n possible results x_j, $j = 1, 2, \ldots, n$, for a measurement of a quantity x, each occurring with its own probability $P(x_j)$, then, since a measurement of x must yield some result in the set $\{x_j\}$, the sum of the probabilities $P(x_j)$ over all possible x_j must be unity:

$$\sum_{j=1}^{n} P(x_j) = 1$$

The quantity x might, for example, be the number of pips on the upper face of a cubical die, with $x_1, x_2, \ldots, x_6$ corresponding to there being 1, 2, ..., 6 pips on that face and $P(x_1), P(x_2), \ldots, P(x_6)$ to the probabilities of there being 1, 2, ..., 6 pips on that face. For these six $P(x_j)$ we must always have

$$\sum_{j=1}^{6} P(x_j) = 1$$

For an unbiased die, each face is equally likely to turn up, and hence all six $P(x_j)$ are $\frac{1}{6}$.

We are, however, more concerned with measurements of continuous quantities that can take on any real value in some range. A length ℓ, for example, can take on any value $\ell > 0$. The distribution of values of such a continuous quantity x is given by a probability density function $p(x)$, where $p(x)\,dx$ is the probability that a measurement of x will yield a value between x and $x + dx$. The statement that the sum of probabilities over all possible values of a continuous variable x must be unity becomes, instead of a discrete sum, an integral over the range of values:

$$\int p(x)\,dx = 1 \tag{25}$$

where the limits on the integration would correspond to the range of possible values of x. The nature of the probability density function of course depends on the quantity being measured, but measurements x subject to (in principle infinitely many) independent random errors are said to be normally distributed, and the probability density function for this normal distribution turns out to be

$$\mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{26}$$

where μ and σ are the mean and standard deviation of the distribution, respectively, and $\exp z$ is an alternate notation for e^z that is convenient when the

expression for the exponent is large. The variable x in the normal distribution ranges from $-\infty$ to $+\infty$, so that eq. (25) becomes [56]

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1 \tag{27}$$

And now the situation that directly concerns us: Suppose we have a set of n independent data points $\{x_i\}$, $i = 1, 2, \ldots, n$, and that we know the true mean μ_x and standard deviation σ_x of the distribution of these data points. This is a hypothetical supposition: in practice, we usually know only sample mean $\widetilde{\mu}_x$ and sample standard deviation $\widetilde{\sigma}_x$ that we calculate from the x_i themselves by eqq. (5) and (6) of p.13 and use as approximations to the true mean and standard deviation.

[56] To verify eq. (27), we need to invoke one of the cheapest tricks in all of mathematics: first we simplify the integral by defining a new variable X by

$$X = \frac{x-\mu}{\sqrt{2\sigma^2}}$$

Then

$$dX = \frac{1}{\sqrt{2\sigma^2}}\, dx \qquad\qquad dx = dX\,\sqrt{2\sigma^2}$$

and the left-hand side of eq. (27) becomes

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-X^2}\, dX\,\sqrt{2\sigma^2} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-X^2}\, dX$$

The trick to doing out this integral is to express its square

$$\left(\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-X^2}\, dX\right)\left(\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-Y^2}\, dY\right) = \frac{1}{\pi}\int_{-\infty}^{\infty} dX \int_{-\infty}^{\infty} dY\, e^{-(X^2+Y^2)}$$

in terms of the polar coordinates $r = \sqrt{X^2+Y^2}$ and θ and note that integrating X and Y from $-\infty$ to $+\infty$ means integrating over the entire plane, that is, $r = 0$ to ∞ and $\theta = 0$ to 2π:

$$dX\, dY = r\, dr\, d\theta = \tfrac{1}{2}\, d(r^2)\, d\theta \qquad\qquad X^2 + Y^2 = r^2$$

$$\begin{aligned}
\frac{1}{\pi}\int_{-\infty}^{\infty} dX \int_{=-\infty}^{\infty} dY\, e^{-(X^2+Y^2)} &= \frac{1}{\pi}\,\frac{1}{2}\int_0^{\infty} d(r^2) \int_0^{2\pi} d\theta\, e^{-r^2} \\
&= \frac{1}{2\pi} \cdot \int_0^{\infty} d(r^2)\, e^{-r^2} \cdot 2\pi \\
&= \frac{1}{2\pi} \cdot \Big[-e^{-r^2}\Big]_0^{\infty} \cdot 2\pi \\
&= \frac{1}{2\pi} \cdot 1 \cdot 2\pi \\
&= 1 \quad \checkmark
\end{aligned}$$

With almost divine foresight, we will define a new variable ξ by

$$\xi = \frac{\widetilde{\mu}_x - \mu_x}{\sigma_x/\sqrt{n}} \tag{28}$$

We will now show that the probability density function $p(\xi)$ of the variable ξ becomes that of the standard normal distribution $\mathcal{N}(0,1)$ as $n \to \infty$, that is, the normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. The first step in establishing this assertion is to rewrite ξ as

$$\xi = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i - \mu_x}{\sigma_x/\sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} \mu_x}{\sigma_x/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{x_i - \mu_x}{\sigma_x} \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i \tag{29}$$

where the y in

$$y_i = \frac{x_i - \mu_x}{\sigma_x} \tag{30}$$

is known as the standardization of x, so called because it is obtained by shifting and rescaling the variable x so that the mean vanishes and the standard deviation is unity:

$$\overline{y} = \frac{\overline{x - \mu_x}}{\sigma_x} = \frac{\overline{x} - \mu_x}{\sigma_x} = \frac{\mu_x - \mu_x}{\sigma_x} = 0$$

and, by eq. (7),

$$\begin{aligned}
\sigma_y^2 &= \overline{y^2} - \overline{y}^2 \\
&= \overline{\left(\frac{x - \mu_x}{\sigma_x} \right)^2} - 0^2 \\
&= \frac{\overline{x^2 - 2\mu_x x + \mu_x^2}}{\sigma_x^2} \\
&= \frac{\overline{x^2} - 2\mu_x \overline{x} + \mu_x^2}{\sigma_x^2} \\
&= \frac{\overline{x^2} - 2\overline{x}\,\overline{x} + \overline{x}^2}{\sigma_x^2} \\
&= \frac{\overline{x^2} - \overline{x}^2}{\sigma_x^2} \\
&= \frac{\sigma_x^2}{\sigma_x^2} \\
&= 1
\end{aligned}$$

The moment generating function $M(z)$ for a probability distribution $p(x)$ is defined as

$$M(z) = \int dx\, p(x)\, e^{xz} \tag{31}$$

M is called the moment generating function because taking the m-th derivative of $M(z)$ with respect to z and setting $z = 0$ will generate the mean $\overline{x^m}$ of x^m, also known as the m-th moment of $p(x)$:

$$\begin{aligned}\left[\frac{d^m}{dz^m}\,M(z)\right]_{z=0} &= \left[\frac{d^m}{dz^m}\int dx\,p(x)\,e^{xz}\right]_{z=0}\\ &= \left[\int dx\,p(x)\,x^m\,e^{xz}\right]_{z=0}\\ &= \int dx\,x^m\,p(x)\\ &= \overline{x^m}\end{aligned} \tag{32}$$

We will show that the variable ξ has a standard normal distribution by showing that it has the same moment generating function as $\mathcal{N}(0,1)$: as you can see from eq. (31), two distributions can have the same moment generating function $M(z)$ for all z only if they have the same probability density function $p(x)$ for all x; if the two $p(x)$ differed, we would be able to find some z for which the two $M(z)$ would not be equal.

For a normal distribution with probability density function

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\!\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

we have, completing the square in the exponent and making the change of variables $u = x - (\mu + \sigma^2 z)$,

$$\begin{aligned}M(z) &= \int_{-\infty}^{\infty} dx\,\frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\!\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,e^{xz}\\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} dx\,\exp\!\left(-\frac{(x-\mu)^2 - 2\sigma^2 zx}{2\sigma^2}\right)\\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} dx\,\exp\!\left(-\frac{x^2 - 2\mu x + \mu^2 - 2\sigma^2 zx}{2\sigma^2}\right)\\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} dx\,\exp\!\left(-\frac{(x-(\mu+\sigma^2 z))^2 - (\mu+\sigma^2 z)^2 + \mu^2}{2\sigma^2}\right)\\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} du\,\exp\!\left(-\frac{u^2 - 2\mu\sigma^2 z - \sigma^4 z^2}{2\sigma^2}\right)\\ &= \exp\!\left(\mu z + \tfrac{1}{2}\sigma^2 z^2\right)\,\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} du\,\exp\!\left(-\frac{u^2}{2\sigma^2}\right)\end{aligned}$$

which, by eq. (27), reduces to

$$M(z) = \exp\!\left(\mu z + \tfrac{1}{2}\sigma^2 z^2\right)$$

And so for the standard normal distribution $\mathcal{N}(0,1)$, for which $\mu = 0$ and $\sigma = 1$, we have

$$M(z) = e^{\frac{1}{2}z^2} \tag{33}$$

We now want to show that our variable ξ has this same moment generating function. First, we note that for sets of n data points x_i all having the same distribution, the moment generating function $M_{\widetilde{\mu}_x}$ of the sample means

$$\widetilde{\mu}_x = \frac{1}{n}\sum_{j=1}^{n} x_j$$

will be simply the product of the moment generating functions for each of the x_i: the x in the e^{xz} in eq. (31) will become $\widetilde{\mu}_x$, and the $p(x)\,dx$ in eq. (31) will become the combined probability, for the set of values $\{x_i\}$, of each x_i taking a value between x_i to $x_i + dx_i$, so that

$$e^{zx} \quad \text{becomes} \quad e^{z\widetilde{\mu}_x} = \exp\left(z\,\frac{1}{n}\sum_{j=1}^{n} x_j\right)$$

$$p(x)\,dx \quad \text{becomes} \quad \prod_{i=1}^{n} dx_i\,p(x_i)$$

The moment generating function $M_{\widetilde{\mu}_x}$ of the sample means will therefore be

$$M_{\widetilde{\mu}_x}(z) = \int\left(\prod_{i=1}^{n} dx_i\,p(x_i)\right)\,\exp\left(z\,\frac{1}{n}\sum_{j=1}^{n} x_j\right) = \prod_{i=1}^{n}\int dx_i\,p(x_i)\,e^{zx_i/n}$$

Comparing this to eq. (31), we see that this is just the product of the moment generating functions M_x for the individual x_i, with z/n in place of z, so that we have

$$M_{\widetilde{\mu}_x}(z) = M_x^n\left(\frac{z}{n}\right)$$

And since, by eq. (29),

$$\xi = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} y_i = \sum_{i=1}^{n}\frac{y_i}{\sqrt{n}}$$

the moment generating function $M_\xi(z)$ for ξ will similarly be

$$M_\xi(z) = M_y^n\left(\frac{z}{\sqrt{n}}\right) \tag{34}$$

If now we use eq. (31) and

$$y_i = \frac{x_i - \mu_x}{\sigma_x}$$

from eq. (30), each $M_y(z/\sqrt{n})$ in eq. (34) further works out to

$$M_y\left(\frac{z}{\sqrt{n}}\right) = \int dy\,p(y)\,e^{yz/\sqrt{n}}$$

$$
\begin{aligned}
&= \int dy\, p(y)\, \exp\left(\frac{x-\mu_x}{\sigma_x}\,\frac{z}{\sqrt{n}}\right) \\
&= \exp\left(-\frac{\mu_x z}{\sigma_x\sqrt{n}}\right)\int dx\, p(x)\, \exp\left(\frac{xz}{\sigma_x\sqrt{n}}\right) \\
&= \exp\left(-\frac{\mu_x z}{\sigma_x\sqrt{n}}\right) M_x\left(\frac{z}{\sigma_x\sqrt{n}}\right)
\end{aligned}
$$

And since, by eq. (32),

$$
\left[\frac{d^m}{dz^m}\, M_x(z)\right]_{z=0} = \overline{x^m}
$$

we have for the first few terms in the Taylor expansion of $M_y(z/\sqrt{n})$

$$
\begin{aligned}
M_y\left(\frac{z}{\sqrt{n}}\right) &= M_y(0) + M_y'(0)\,\frac{z}{\sqrt{n}} + \tfrac{1}{2}M_y''(0)\left(\frac{z}{\sqrt{n}}\right)^2 + \cdots \\
&\approx 1 + \overline{y}\,\frac{z}{\sqrt{n}} + \tfrac{1}{2}\,\overline{y^2}\,\frac{z^2}{n}
\end{aligned}
$$

where we have dropped the cubic and higher terms, which can, with a more careful analysis, be shown to vanish in the limit in which we are interested, that is, as $n \to \infty$. Since y is the standardization of x, it has, as shown following eq. (30), vanishing mean and unit standard deviation, so that the general relation $\sigma_y^2 = \overline{y^2} - \overline{y}^2$ of eq. (7) yields

$$
1 = \overline{y^2} - 0^2 \qquad\qquad \overline{y^2} = 1
$$

and the Taylor expansion further reduces to

$$
M_y\left(\frac{z}{\sqrt{n}}\right) \approx 1 + \frac{z^2}{2n}
$$

Thus eq. (34) becomes

$$
M_\xi(z) = M_y^n\left(\frac{z}{\sqrt{n}}\right) \approx \left(1 + \frac{z^2}{2n}\right)^n
$$

We are, again, interested in the limit $n \to \infty$, which allows us to invoke the general mathematical result [57]

$$
\lim_{n\to\infty}\left(1 + \frac{u}{n}\right)^n = e^u
$$

[57] If you have forgotten this relation, a quick and dirty way to see it is by doing a binomial expansion:

$$
\begin{aligned}
\lim_{n\to\infty}\left(1 + \frac{u}{n}\right)^n &= \lim_{n\to\infty}\sum_{k=0}^{n} {}_nC_k\, 1^{n-k}\left(\frac{u}{n}\right)^k \\
&= \lim_{n\to\infty}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\left(\frac{u}{n}\right)^k
\end{aligned}
$$

Thus

$$\lim_{n\to\infty} M_\xi(z) = \lim_{n\to\infty}\left(1+\frac{z^2}{2n}\right)^n = e^{\frac{1}{2}z^2}$$

Comparing this result to eq. (33), we see that as the number of data points becomes infinite, the moment generating function for ξ becomes the moment generating function for the normal distribution $\mathcal{N}(0,1)$ and hence that the distribution of ξ for a large number n of data points becomes $\mathcal{N}(0,1)$ — the proposition we had originally set out to prove. If we further compare the definition (28) of ξ,

$$\xi = \frac{\widetilde{\mu}_x - \mu_x}{\sigma_x/\sqrt{n}}$$

to that of y in eq. (30),

$$y = \frac{x-\mu_x}{\sigma_x}$$

we see that this means that the sample mean $\widetilde{\mu}_x$ tends to the true mean μ_x with a standard deviation $\sigma_x/\sqrt{n}$. That is, the standard deviation of the mean for a set of n measurements x_i having standard deviation σ_x approaches $\sigma_x/\sqrt{n}$ when n is large. This is the reason why standard error is taken to be the standard deviation divided by the square root of the number of measurements.

This result also helps us see why independent random experimental errors tend to produce normally distributed measurements x_i: The central limit theorem tells us that the distribution of ξ, and hence of the sample means of sets $\{x_i\}$ of data, is normal. In a sense, the theorem can be regarded as the statement that the distribution of distributions is a normal distribution: as multiple sets of data $\{x_i\}$ become infinitely numerous, their sample means not only approach the true mean, but the distribution of these sample means approaches the normal distribution. In the special case that each data set consists of a single measurement, the sample mean is the value of that single measurement, and the distribution of the values of the measurements themselves is therefore also normal.

One can, for example, think of a random error as the flip of a coin: just as the outcome of a flip of a coin is either heads or tails, a random error results in either an increase or a decrease in the measured value. Coin flips obey the binomial distribution, so called because it is generated in the same way as the powers of a binomial. If the probability of heads is p, then the probability q of tails is $q = 1-p$. If we flip a coin n times, there is a one-to-one correspondence

$$\begin{aligned}
&= \lim_{n\to\infty}\sum_{k=0}^{n}\frac{n!}{n^k(n-k)!}\frac{u^k}{k!}\\
&= \lim_{n\to\infty}\sum_{k=0}^{\infty}\frac{u^k}{k!}\\
&= e^u
\end{aligned}$$

between the 2^n possible outcomes of these flips and the 2^n terms of $(p+q)^n$: if we expand $(p+q)^n$ into its factors,

$$(p+q)^n = (p+q)(p+q)(p+q)\cdots(p+q)$$

we see that expanding the product algebraically involves picking one of either p or q from each of the n factors $(p+q)$, just as enumerating the possible outcome of flipping a coin n times involves picking one of either heads (with probability p) or tails (with probability q) from each flip. From the binomial theorem

$$(p+q)^n = \sum_{k=0}^{n} {}_nC_k p^k q^{n-k} = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!}\, p^k q^{n-k}$$

we therefore see that the probability of getting k heads in n flips is

$$P(k) = {}_nC_k p^k q^{n-k} = \frac{n!}{k!(n-k)!}\, p^k q^{n-k}$$

The mean number μ of heads in n flips will be the average of the various possible numbers of heads, each weighted by its probability of occurrence: if numerically we regard a head as unity and a tail as zero, the value associated with k heads is $k \cdot 1 = k$ and

$$\mu = \sum_{k=0}^{n} kP(k) = \sum_{k=0}^{n} k\, \frac{n!}{k!(n-k)!}\, p^k q^{n-k}$$

which, since the overall factor of k kills the contribution from the $k=0$ term, reduces to

$$\begin{aligned}
\mu &= \sum_{k=1}^{n} k\, \frac{n!}{k!(n-k)!}\, p^k q^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}\, p^k q^{n-k} \\
&= \sum_{k=1}^{n} \frac{n\cdot(n-1)!}{(k-1)!\Big((n-1)-(k-1)\Big)!}\, (p\cdot p^{k-1}) q^{(n-1)-(k-1)} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\Big((n-1)-(k-1)\Big)!}\, p^{k-1} q^{(n-1)-(k-1)}
\end{aligned}$$

If now we rewrite this sum in terms of a new index $k' = k-1$, the range $1 \le k \le n$ will become $0 \le k' \le n-1$:

$$\begin{aligned}
\mu &= np \sum_{k'=0}^{n-1} \frac{(n-1)!}{k'!\Big((n-1)-k'\Big)!}\, p^{k'} q^{(n-1)-k'} \\
&= np(p+q)^{n-1} \\
&= np\Big(p+(1-p)\Big)^{n-1}
\end{aligned}$$

$$= np \cdot 1^{n-1}$$
$$= np$$

With some similar gerrymandering, we also have

$$\begin{aligned}
\overline{k^2} &= \sum_{k=0}^{n} k^2 P(k) \\
&= \sum_{k=0}^{n} k^2 \, \frac{n!}{k!(n-k)!} \, p^k q^{n-k} \\
&= \sum_{k=1}^{n} k^2 \, \frac{n!}{k!(n-k)!} \, p^k q^{n-k} \\
&= \sum_{k=1}^{n} k \, \frac{n!}{(k-1)!(n-k)!} \, p^k q^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \, (kp^k) q^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \left(p \, \frac{\partial}{\partial p} \, p^k \right) q^{n-k} \\
&= p \, \frac{\partial}{\partial p} \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \, p^k q^{n-k} \\
&= p \, \frac{\partial}{\partial p} \Big(np(p+q)^{n-1} \Big) \\
&= p \Big(n(p+q)^{n-1} + np \cdot (n-1)(p+q)^{n-2} \Big) \\
&= p \Big(n \cdot 1^{n-1} + np \cdot (n-1) \cdot 1^{n-2} \Big) \\
&= np(1 + p(n-1))
\end{aligned}$$

Using eq. (7), we thus obtain for the square of the standard deviation

$$\sigma^2 = \overline{k^2} - \overline{k}^2 = np(1 + p(n-1)) - \mu^2 = np(1 + p(n-1)) - (np)^2$$

which, when we do out the algebra, yields

$$\sigma^2 = np(1-p) = npq \qquad\qquad \sigma = \sqrt{npq}$$

The average number of heads per flip is thus

$$\frac{\mu}{n} = \frac{np}{n} = p$$

More significantly in our current context, the standard deviation per flip is

$$\frac{\sigma}{n} = \frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}} \sim \frac{1}{\sqrt{n}}$$

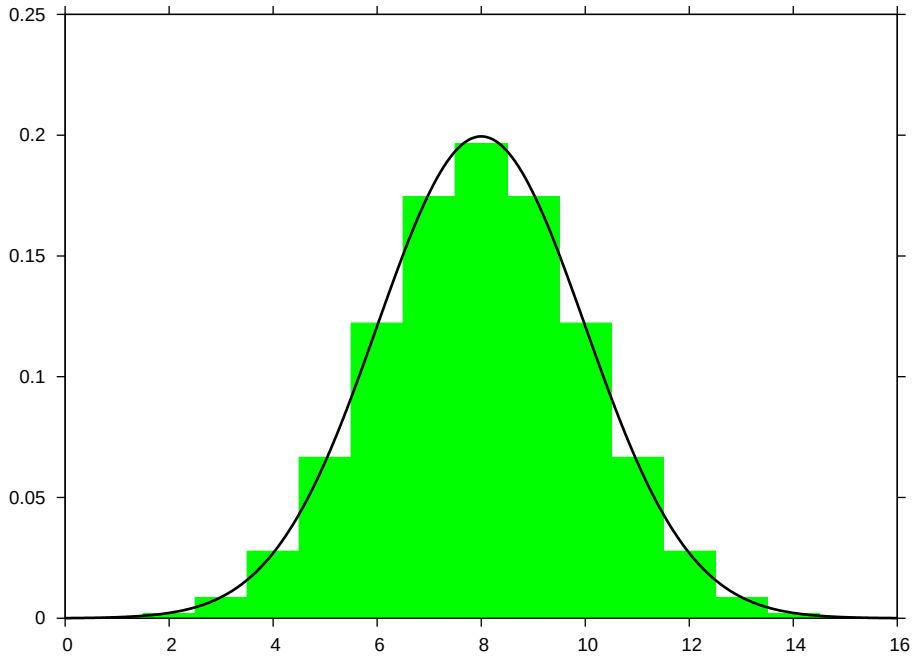

Figure 6: Sixteen is Bigger Than You Thought

In fact, we expect that as the number n of flips approaches infinity, the binomial distribution will reproduce the normal distribution. Fig. (6) shows the case of sixteen flips of an unbiased coin — a coin for which $p = q = \frac{1}{2}$ and thus $\mu = np = \frac{1}{2}n$ and $\sigma = \sqrt{npq} = \frac{1}{2}\sqrt{n}$. The black curve is the plot of the normal distribution with $\mu = \frac{1}{2}(16) = 8$ and $\sigma = \frac{1}{2}/\sqrt{16} = \frac{1}{8}$, and the green boxes are the probabilities of the number of heads indicated on the horizontal axis. As you can see from the figure, even though sixteen is considerably less than infinity, the normal and binomial distributions already agree very closely.

C The Standard Deviation Ain't the Average Distance from the Mean

While one can conjure up distributions for which the standard deviation does, by fluke, equal the average distance from the mean, in general the standard deviation and average distance are not the same, and we will now demonstrate their inequality for the normal distribution. We will assume that you already have some — though not much — familiarity with probability density functions and the normal distribution, either from a previous life in math or from the first few pages of Appendix B. Those of you with greater familiarity may find that you can skip over our derivations of some preliminary results in the following subsections and cut right to the chase on p.69.

From eq. (26) of Appendix B, the probability density function $\mathcal{N}$ for a normal distribution of mean μ and standard deviation σ is given by

$$\mathcal{N}(\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Derivation of the Normalization Factor $\frac{1}{\sqrt{2\pi\sigma^2}}$

If you are familiar with the normal distribution, you know that the normalization factor

$$\frac{1}{\sqrt{2\pi\sigma^2}}$$

does indeed normalize the integral of this probability density function over all values of x to unity:

$$\int_{-\infty}^{\infty} dx\,\frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = 1 \tag{35}$$

If this is news to you, we will now repeat the proof from footnote 56 on p.57: first we simplify the integral by defining a new variable X by

$$X = \frac{x-\mu}{\sqrt{2\sigma^2}} \tag{36}$$

Then

$$dX = \frac{1}{\sqrt{2\sigma^2}}\,dx \qquad\qquad dx = dX\,\sqrt{2\sigma^2}$$

and the left-hand side of eq. (35) becomes

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-X^2}\,dX\,\sqrt{2\sigma^2} = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-X^2}\,dX$$

The trick to doing out this integral is to express its square

$$\left(\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-X^2}\,dX\right)\left(\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-Y^2}\,dY\right) = \frac{1}{\pi}\int_{-\infty}^{\infty} dX\int_{-\infty}^{\infty} dY\,e^{-(X^2+Y^2)}$$

in terms of the polar coordinates $r = \sqrt{X^2+Y^2}$ and θ and note that integrating X and Y from $-\infty$ to $+\infty$ means integrating over the entire plane, that is, $r = 0$ to ∞ and $\theta = 0$ to 2π:

$$dX\,dY = r\,dr\,d\theta = \tfrac{1}{2}\,d(r^2)\,d\theta \qquad\qquad X^2+Y^2 = r^2$$

$$\begin{aligned}
\frac{1}{\pi}\int_{-\infty}^{\infty} dX \int_{=-\infty}^{\infty} dY\, e^{-(X^2+Y^2)} &= \frac{1}{\pi}\frac{1}{2}\int_0^{\infty} d(r^2)\int_0^{2\pi} d\theta\, e^{-r^2} \\
&= \frac{1}{2\pi}\cdot\int_0^{\infty} d(r^2)\, e^{-r^2}\cdot 2\pi \\
&= \frac{1}{2\pi}\cdot\left[-e^{-r^2}\right]_0^{\infty}\cdot 2\pi \\
&= \frac{1}{2\pi}\cdot 1\cdot 2\pi \\
&= 1 \quad \checkmark
\end{aligned}$$

The mathematical result

$$\int_{-\infty}^{\infty} e^{-X^2}\, dX = \sqrt{\pi} \tag{37}$$

that we have just obtained is in fact much more generally useful than you would think from this limited context.[58]

Proofs that μ and σ are the Mean and Standard Deviation

If you were not already familiar with the derivation of the normalization of the normal distribution, you probably also have not verified that its mean and standard deviation are indeed μ and σ, so we might as well establish this while we're at it. We wouldn't want your education to be incomplete.

The mean $\overline{x}$ of a distribution with probability density function $p(x)$ is given by integrating x over all of its possible values, each weighted by the probability density function:

$$\overline{x} = \int dx\, x\, p(x)$$

For the normal distribution we therefore have

$$\overline{x} = \int_{-\infty}^{\infty} dx\, x\, \frac{1}{\sqrt{2\pi\sigma^2}}\, \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

With the same change of variables that we used above in eq. (36),

$$X = \frac{x-\mu}{\sqrt{2\sigma^2}} \qquad\qquad dX = \frac{1}{\sqrt{2\sigma^2}}\, dx$$

$$x = \mu + X\sqrt{2\sigma^2} \qquad\qquad dx = dX\,\sqrt{2\sigma^2}$$

[58] Also more frequently useful than you would think are the kindred integrals

$$\int_0^{\infty} X^{2n+1} e^{-X^2}\, dX = \tfrac{1}{2}n! \qquad\qquad \int_0^{\infty} X^{2n} e^{-X^2}\, dX = \frac{(2n-1)!!}{2^{n+1}}\sqrt{\pi}$$

These can be derived using integration by parts in conjunction with eq. (37) or with the change of variables $u = X^2$, which we will leave as an exercise for those readers who are interested.

we have

$$\begin{aligned}\overline{x} &= \int_{-\infty}^{\infty} (dX\,\sqrt{2\sigma^2})\,(\mu + X\sqrt{2\sigma^2})\,\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-X^2} \\ &= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} dX\,(\mu + X\sqrt{2\sigma^2})\,e^{-X^2} \\ &= \frac{\mu}{\sqrt{\pi}}\int_{-\infty}^{\infty} dX\,e^{-X^2} + \sqrt{\frac{2\sigma^2}{\pi}}\int_{-\infty}^{\infty} dX\,Xe^{-X^2}\end{aligned}$$

From our result eq. (37), the first of these integrals is just $\sqrt{\pi}$, and the second, its integrand being an odd function of X, will vanish when integrated from $X = -\infty$ to $X = +\infty$. We therefore arrive at

$$\overline{x} = \frac{\mu}{\sqrt{\pi}}\sqrt{\pi} + 0 = \mu \quad \checkmark$$

Recall that the standard deviation is, by definition, the root-mean-square deviation from the mean:

$$\sqrt{\overline{(x-\mu)^2}}$$

For a distribution with probability density function $p(x)$, the mean value of $(x-\mu)^2$ (the square root of which is the standard deviation we seek) is given by the integral of $(x-\mu)^2$ over all values of x, each weighted by the probability density function:

$$\overline{(x-\mu)^2} = \int dx\,(x-\mu)^2\,p(x)$$

For the normal distribution we therefore have

$$\overline{(x-\mu)^2} = \int_{-\infty}^{\infty} dx\,(x-\mu)^2\,\frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

With the same change of variables

$$\begin{aligned} X &= \frac{x-\mu}{\sqrt{2\sigma^2}} & \qquad dX &= \frac{1}{\sqrt{2\sigma^2}}\,dx \\ x &= \mu + X\sqrt{2\sigma^2} & \qquad dx &= dX\,\sqrt{2\sigma^2}\end{aligned}$$

yet again, we have

$$x - \mu = X\sqrt{2\sigma^2} = X\sqrt{2\sigma^2}$$

and

$$\begin{aligned}\overline{(x-\mu)^2} &= \int_{-\infty}^{\infty}\left(dX\,\sqrt{2\sigma^2}\right)\left(X\sqrt{2\sigma^2}\right)^2\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-X^2} \\ &= \frac{2\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{\infty} dX\,X^2e^{-X^2}\end{aligned}$$

We can carry out this integration by parts: if we apply

$$\int u\,dv = uv - \int v\,du$$

with

$$\begin{aligned} u &= X & du &= dX \\ dv &= dX\, Xe^{-X^2} = -\tfrac{1}{2}\, d(e^{-X^2}) & v &= -\tfrac{1}{2}e^{-X^2} \end{aligned}$$

we obtain

$$\begin{aligned} \int_{-\infty}^{\infty} dX\, X^2 e^{-X^2} &= \int_{-\infty}^{\infty} (X)(dX\, Xe^{-X^2}) \\ &= (X)(-\tfrac{1}{2}e^{-X^2})\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-\tfrac{1}{2}e^{-X^2})\, dX \\ &= 0 + \tfrac{1}{2}\int_{-\infty}^{\infty} dX\, e^{-X^2} \\ &= \tfrac{1}{2}\sqrt{\pi} \end{aligned}$$

We therefore arrive at

$$\overline{(x-\mu)^2} = \frac{2\sigma^2}{\sqrt{\pi}}\,\frac{\sqrt{\pi}}{2} = \sigma^2$$

and

$$\sqrt{\overline{(x-\mu)^2}} = \sigma \quad \checkmark$$

Average Distance from the Mean

The average distance from the mean will be the average of $|x-\mu|$ over all values of x, which for a distribution with probability density function $p(x)$ will be given by

$$\overline{|x-\mu|} = \int dx\, |x-\mu|\, p(x)$$

For the normal distribution we therefore have

$$\overline{x} = \int_{-\infty}^{\infty} dx\, |x-\mu|\, \frac{1}{\sqrt{2\pi\sigma^2}}\, \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Once again we use the same change of variables that we did in eq. (36) and in the preceding subsections:

$$\begin{aligned} X &= \frac{x-\mu}{\sqrt{2\sigma^2}} & dX &= \frac{1}{\sqrt{2\sigma^2}}\, dx \\ x &= \mu + X\sqrt{2\sigma^2} & dx &= dX\, \sqrt{2\sigma^2} \end{aligned}$$

We thus have

$$|x-\mu| = |X\sqrt{2\sigma^2}| = |X|\sqrt{2\sigma^2}$$

and

$$\overline{|x-\mu|} = \int_{-\infty}^{\infty} (dX\, \sqrt{2\sigma^2})\, |X|\sqrt{2\sigma^2}\, \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-X^2}$$

$$= \sqrt{\frac{2\sigma^2}{\pi}} \int_{-\infty}^{\infty} dX \, |X| \, e^{-X^2}$$

Since the integrand is an even function of X, the result of integrating from $X = -\infty$ to $X = +\infty$ will be just double the result of integrating from $X = 0$ to $X = +\infty$, over which range $|X| = X$. We therefore arrive at

$$\begin{aligned}
\overline{|x-\mu|} &= 2\sqrt{\frac{2\sigma^2}{\pi}} \int_0^{\infty} dX \, X e^{-X^2} \\
&= 2\sqrt{\frac{2\sigma^2}{\pi}} \int_0^{\infty} \tfrac{1}{2} \, d(X^2) \, e^{-X^2} \\
&= 2\sqrt{\frac{2\sigma^2}{\pi}} \int_0^{\infty} \tfrac{1}{2} \, du \, e^{-u} \\
&= 2\sqrt{\frac{2\sigma^2}{\pi}} \left[-\tfrac{1}{2} e^{-u} \right]_0^{\infty} \\
&= 2\sqrt{\frac{2\sigma^2}{\pi}} \cdot \frac{1}{2} \\
&= \sigma\sqrt{\frac{2}{\pi}}
\end{aligned}$$

Were the standard deviation the average distance from the mean, we would have arrived at $\overline{|x-\mu|} = \sigma$, whereas we actually have $\sigma\sqrt{2/\pi}$. And whatever else you may say about $\sqrt{2/\pi}$, it's definitely not equal to 1:

$$\sqrt{\frac{2}{\pi}} \approx 0.797884560803$$

Since everyone always enjoys pictures, especially color pictures, let us introduce the standard normal distribution, which is the normal distribution

$$\mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

shifted so that the mean $\mu = 0$ and scaled so that the standard deviation $\sigma = 1$:[59]

$$\mathcal{N}(0, 1) = \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}x^2}$$

[59] Note that while we can set $\mu = 0$ without worrying about units, setting $\sigma = 1$ is equivalent to specifying a system of units: we are now measuring x in units of σ. If, for example, our x variable was a length and σ was 4 cm, then we are changing to a length scale where 1 (meaning 1 unit) $\leftrightarrow$ 4 cm, 2 $\leftrightarrow$ 8 cm, 3 $\leftrightarrow$ 12 cm, etc.

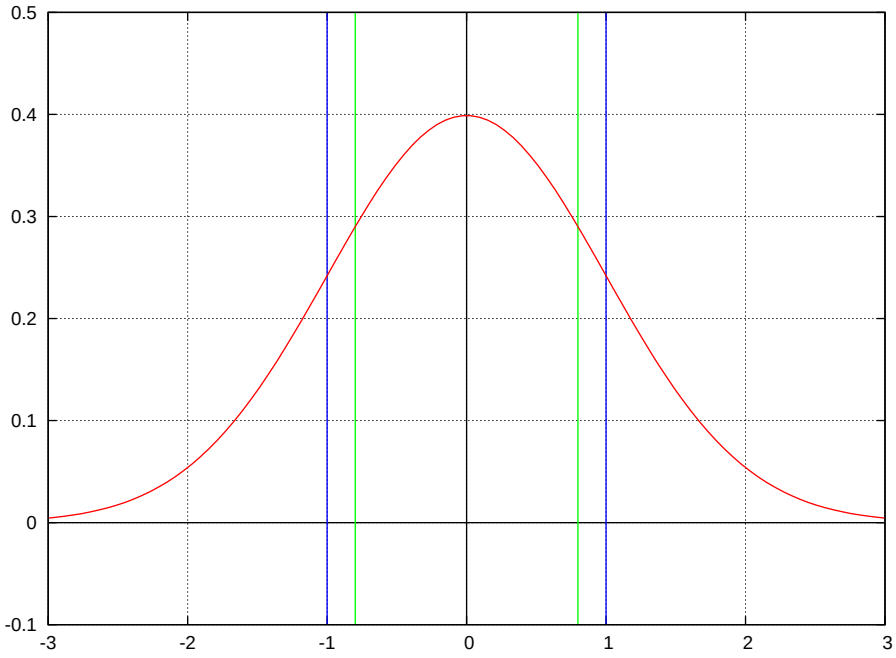

Figure 7: The Standard Normal Distribution

Fig. (7) shows a plot of the standard normal distribution, with blue lines marking a distance of 1σ to either side of the mean and green lines marking the average distance $\sqrt{2/\pi}$ from the mean. So there.